An Application of the James-Stein Estimation Method in Modeling of Mortality Rates

by

Jinjin (Summer) Shan

B.Sc., University of British Columbia, 2019

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2023

Declaration of Committee

Name:	**Jinjin (Summer) Shan**
Degree:	**Master of Science**
Thesis title:	**An Application of the James-Stein Estimation Method in Modeling of Mortality Rates**
Committee:	**Chair:** Liangliang Wang Associate Professor, Statistics and Actuarial Science
	Cary Chi-Liang Tsai Supervisor Professor, Statistics and Actuarial Science
	Himchan Jeong Committee Member Assistant Professor, Statistics and Actuarial Science
	Haolun Shi Examiner Assistant Professor, Statistics and Actuarial Science

Abstract

It is known that the James-Stein estimation method outperforms the Maximum Likelihood Estimation method when we estimate a $p-$dimension independently distributed random variable with $p \geq 3$. In this project, an explicit formula based on a modified James-Stein estimation is first derived to forecast a $p-$dimension random variable. Then the modified James-Stein estimator is applied to forecasting of mortality rates for 10, 20 and 30 years for six populations (both genders of the U.S., the U.K., and Japan). Moreover, some underlying mortality models (the Lee-Carter model, the Cairns-Blake-Dowd model, the M6 and M7 models, and the Renshaw-Haberman model) are also used in the forecasting of mortality rates to compare their forecast performances with the modified James-Stein estimation method. The results show that the modified James-Stein estimation method has the lowest overall average estimation error compared to all other mortality models. Finally, the shrinkage effect of the modified James-Stein estimate is studied with numerical illustrations for six populations and three forecasting years.

Keywords: James-Stein estimator; Shrinkage; Mortality model; Lee-Carter model; CBD model

Acknowledgements

Deciding to pursue a Master's degree in Actuarial Science at the Simon Fraser University is one of the best decisions I have ever made, and Dr. Cary Chi-Liang Tsai's continuous support and encouragement guided me through the program course work and the degree project. Thank you, Professor Tsai for leading me into the actuarial world and providing valuable knowledge and suggestions that fuelled this degree project.

I would like to thank Dr. Liangliang Wang for serving as the chair, Dr. Himchan Jeong for serving as the committee member, and Dr. Haolun Shi for serving as the examiner for my degree project defence. I also would like to thank SFU's Co-operative Education Department that provided valuable opportunities for me to intern at Sun Life, Intact, and Milliman to apply learnings in the workplace effectively. Finally, I have to give a big hug to my families and friends that support me and wipe away my tears when the degree project gets too difficult. Cheers!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Mortality rate is one of the key factors in determining the premiums and reserves of life insurance and annuity products. Modeling mortality rates is important to life insurers and annuity providers because the forecasting results directly affect their profitability and solvency. Therefore, developing an accurate and effective mortality model can help life insurers and annuity providers better manage mortality risk from life insurance (insureds die earlier than expected) and longevity risk from annuity products (annuitants live longer than expected), respectively.

The Lee-Carter [7] and Cairns-Blake-Dowd [5] models are the two most widely used and cited mortality models in actuarial literature. The LC model is highly cited due to its simple decomposition nature such that the logarithm of central death rate is split into a mortality trend and two age-specific mortality parameters, and the CBD model gains its popularity due to its excellence in predicting mortality rates for seniors. However, the two models have drawbacks. For instance, the LC model forecasts mortality rates for seniors poorly, and the CBD model does not give good predictions of the mortality rates for younger people. More discussions on the extensions of these two mortality models and other models are covered in Chapter 2.

The James-Stein (JS) estimator is a biased estimator that achieves an outstanding performance when estimating the means of three or more random variables, and it is proposed by Stein [19] and later improved by James and Stein [13]. Maximum likelihood estimation and Bayesian estimation are the two most popular statistical inference methods, but the JS estimation method uses the empirical Bayes approach and achieves admissibility under high-dimensional mean estimation. In other word, there is no other estimation method that is always better than the JS method for a high number of random variables, and this finding has shocked the statistical community when the proof is proposed by James and Stein [14].

This project applies a modification of the JS estimation method to mortality rate forecasting to demonstrate its superiority compared to some underlying mortality models.

1.2 Motivation

To develop a mortality model, We are given a mortality data set composed of mortality rates indexed by age and year. Since a mortality rate for age x and year t is treated as a random variable, modeling mortality rates with a given mortality data set involves an estimation problem in two-dimensional random variables. Tsai and Lin [20] propose a mortality model with a very good forecasting performance, based on non-parametric credibility estimator, where the credibility estimator for age x is a weighted average of the sample mean for that age and the true mean, where the true mean can be estimated by the average of the sample means for all ages. We find that a modification of James-Stein estimator has the same function form as the credibility estimator, where the modified JS estimator (a shrinkage estimator) for an age, under a mortality rate forecasting context, is a weighted mortality rate estimate of the sample mean over years for that age and the overall mean of the entire mortality data.

Therefore, the original James-Stein estimator is modified and applied to the mortality rate forecasting in this project. The ultimate goal of this project aims to use the modified James-Stein estimation method to forecast mortality rates for a range of ages, and show that the modified James-Stein estimator outperforms the underlying mortality models, including the Lee-Carter (LC) model, the Cairns-Blake-Dowd (CBD) model, the M6 and M7 models (two extensions of the CBD model), and the Renshaw-Haberman (RH) model.

1.3 Outline

This project starts with a literature review on some of the popular mortality models, as well as the formulation of the James-Stein estimation method in Chapter 2. Next, Chapter 3 provides a detailed mathematical derivation of the JS method and a modification of the JS method used under a mortality rate forecasting context. Chapter 4 gives numerical illustrations in the application of the modified JS estimator to forecasting of mortality rates for three countries, and compares the forecasting performance of the JS model with some underlying popular mortality models, which are evaluated by three error metrics. Moreover, we define and calculate the shrinkage amount of the modified JS estimates. Finally, the project is finished with a conclusion in Chapter 5.

Chapter 2

Literature Review

2.1 Mortality Models

Mortality forecasting is a widely studied topic for pricing and reserving of life insurance and annuity products, which need mortality rates for a wide range of years to calculate the expected payout in each future year. Under the actuarial equivalent principle, the present value of all expected payouts should equal to the present value of all premiums collected from the insureds. Furthermore, the death benefit is a big lump sum of money payable to the beneficiary when the insured passes away. An accurate mortality model can help the life insurer set up a sufficient amount of reserve to prepare for the future payout. Therefore, building an accurate mortality forecasting is essential to life insurers. This section reviews a good deal of mortality forecasting models that are widely used, along with the advantages and weaknesses of these models.

The first model is the Lee-Carter model, proposed by Lee and Carter [7], which forecasts the natural logarithm of the central death rate for a range of ages. It is the most popular and cited mortality model in actuarial literature, where the natural logarithm of the central death rate is modeled by a mortality time trend and two age-specific mortality parameters which can be estimated through singular value decomposition (SVD). However, the drawback of the Lee-Carter model is that it often under-estimates the life expectancy at older ages due to the insensitivity of the natural logarithm of the central death rate at each of age-specific mortality parameters [1]. The Lee-Carter model is further improved by the RH model, proposed by Renshaw and Haberman [17], to capture similarities within cohorts and enhance the mortality predictions by adding a cohort factor to the LC model.

The second widely cited mortality model is the Cairns-Blake-Dowd (CBD) model that forecasts the logit function of the one-year death probability with an intercept and a slope as two time trend parameters [5]. The two time trend parameters can be estimated by the simple least square method. Furthermore, the slope time trend parameter provides exponential dynamics for older ages and makes the CBD model an excellent mortality

model for longevity insurance products [5]. Thus, the CBD model may not be the best model to forecast the mortality rates for younger ages.

Next, the CBD model is further extended to the M6 and M7 models. The M6 model includes a cohort effect term that is adopted from the RH model to improve the mortality rate estimations for younger ages on top of excellent older age mortality predictions by the CBD model [6]. The M7 model is also an extension to the CBD model that has a quadratic term in addition to the cohort effect term [6]. The quadratic term captures the potential curvature in the historic logit function of one-year death probability.

Finally, Tsai and Yang [22] propose the linear relational (LR) model that utilize the relationship between a target mortality sequence and a base mortality sequence to predict future mortality rates. The intercept and slope parameters for the two time trend parameters in the LR model can be estimated through the simple linear regression or the random walk (RW) with drift, and they are called the LR-LR model and the LR-RW model, respectively. Moreover, Tsai and Lin [21] incorporate the Bühlmann credibility theory into the LC, CBD, and LR mortality models to further improve their performances in mortality prediction. See Appendix *A* for the detailed formulas for each of the LC, CBD, RH, M6 and M7 mortality models.

2.2 Non-Actuarial Applications of the James-Stein Estimation Method

James and Stein [14] propose a biased estimator for the mean of high-dimension random vectors, called James-Stein estimator (a shrinkage estimator), which performs better than the maximum likelihood estimator when the dimension of the random vector is more than or equal to 3. The James-Stein estimation method has been applied to a wide range of topics and achieved great results. The followings are three applications of the James-Stein method to high-dimension estimation problems.

Hausser and Strimmer [12] use a modified James-Stein shrinkage estimator to estimate the entropy values in gene expressions, which estimates the amount of information carried by the gene. The shrinkage estimator utilizes a weighted average of the shrinkage target and the maximum likelihood estimate to produce the James-Stein estimator. The simulation results show that the James-Stein estimator achieves the lowest mean squared errors among the other nine proposed estimators, including maximum likelihood estimator, Miller-Madow estimator, Bayesian estimator, NSB estimator, and Chao-Shen estimator. Moreover, the James-Stein estimator performs surprisingly well when the sample size is small, whereas the other proposed estimators deteriorate drastically as the sample size shrinks. Thus, the modified James-Stein estimator is proved to be an efficient and easy-to-compute method to work well for high-dimension and large scale estimation problems.

Jorion [15] applies the Bayes-Stein estimator to portfolio selection problems, such that Stein's estimator is embedded within the empirical Bayes' posterior distribution to estimate stock returns. The simulation results show that the Bayes-Stein estimator achieves the lowest empirical risk compared to the other proposed estimators, including certainty equivalence, Bayes diffuse prior, and minimum variance estimation methods. It's because the Bayes-Stein estimator puts more weight on the overall sample mean than the individual sample means when the sample size is small and the variation of stock returns is large. Thus, the shrinkage behavior of the Bayes-Stein estimator improves estimation accuracy.

Finally, Efron and Hastie [9] use the James-Stein method to predict the score success rate of eighteen baseball players in a tournament, and compare the predictions between the James-Stein method and the maximum likelihood estimation method against the true observations. The result again shows that the sum of squared errors from the James-Stein method is a half of that from the maximum likelihood estimation method. Thus, the James-Stein estimation method has achieved great successes in the non-actuarial applications.

2.3 Actuarial Applications of the James-Stein Estimation Method

There are few past studies on actuarial applications of the James-Stein estimation method. Marshall [16] applies the James-Stein shrinkage estimator to estimation of infant mortality rates in Auckland, New Zealand. A modified empirical Bayes James-Stein type estimator achieves a lower estimated standard error than the maximum likelihood estimator through shrinkage, such that the regions with high mortality rate estimates are shrunk towards the mean and the regions with low mortality rate estimates are brought up towards the mean. Moreover, the study compares local and global James-Stein type estimators, where the estimated mortality rates are shrunk toward a local neighborhood mean or the overall mean, respectively. The main goal of Marshall [16] is to estimate human mortality rates due to different diseases on a map. Thus, the disease simulation study shows that the local James-Stein estimator performs better when the disease is common but has different spread pattern geographically, whereas the global James-Stein estimator works well when the disease is rare but affects a large region uniformly.

The applications, with the James-Stein estimation method, in Marshall [16] intend to create mortality maps resulting from diseases, whereas this project intends to modify the James-Stein estimator to forecast mortality rates for a wide range of ages. Furthermore, this project also has a goal to develop an effective mortality model which will benefit life insurers and annuity providers. Therefore, this project is the first article to apply the James-Stein estimation method to the human mortality data, and expect to achieve better forecasting results than the other underlying mortality models.

2.4 Credibility Theory Inspiration

Credibility theory is consolidated by Braverman [3] to help casualty actuaries weight historical and observed loss data for a fair insurance premium rate. The classical credibility formula takes the following form,

$$RP = Z\,D + (1 - Z)\,M,$$

where the credibility premium equals to the weight Z (credibility factor) times the observed loss data D plus the weight $(1-Z)$ times the historical manual loss value M. The credibility factor Z acts like a regulation factor to control the amount of influences that the new set of observations bring. Bühlmann [4] proposes a parametric Bühlmann credibility model, where the credibility factor Z is calculated as the number of claims divided by the sum of the quotient and the number of claims, and the quotient equals the expected value of the process variance over the variance of the hypothetical mean; the observed loss data D is replaced by the sample mean of past claims; and the historical manual loss value M is replaced by the true mean of claims.

Although credibility theory is mainly used in casualty insurances, it can apply to the life insurance. For instance, Tsai and Lin [21] incorporate Bühlmann credibility to the Lee-Carter model, the Cairns-Blake-Dowd model, and a linear relational (LR) model to forecast mortality rates and achieve lower estimation errors than without incorporation of the credibility theory. Recently, Bozikas [2] uses the credibility theory framework to price life annuity products when the historical mortality data information is scarce, and achieves great results with Greece mortality data. Bozikas [2] also proposes a concern on the current mortality rate estimation methods used by insurance companies and government agencies. It gives longevity risk to annuity products because people are living longer nowadays than past. Credibility theory framework updates the mortality rate estimates continuously through the credibility factor to have a more up-to-date mortality rate forecast to better price life insurance and annuity products.

Thus, the credibility theory can apply to not only casualty insurance but also life insurance through pooling. The James-Stein estimator will be explained further in details in the next chapter, and the proposed modified James-Stein estimator has an explicit form which is very alike to the above credibility formula, and thus has similar implications, such that the mean of mortality data for an age is pooled as the observed data D for that age, and the overall mean of the entire mortality data for all ages is similar to the historical manual value M. Therefore, the application of the James-Stein estimation to mortality rate estimation is inspired by the credibility theory.

Chapter 3

Models and Methods

In this chapter, the James-Stein estimator is introduced along with the Stein's paradox and derivation of the James-Stein estimator in an explicit form. Moreover, a multivariate form of the James-Stein estimator is presented to ease the estimations in the future mortality rates across different ages. Lastly, the application of the James-Stein estimation in estimating the future mortality rates is described in details.

3.1 James-Stein Estimator

3.1.1 Introduction to the James-Stein Estimator

When estimating the mean of a population, it is natural to use the average of the sampled data. For example, the mean of a population can be estimated through taking the mean of a sample collected. This simple averaging method is also called the maximum likelihood estimation (MLE) method when the distribution of the sample measurements is approximately normal, because the probability that the sample mean equals the true mean is maximized at the average value. Therefore, the MLE method is the gold standard and proved to be an admissible estimator, such that there is no other estimator better than it when estimating population averages [18]. The James-Stein (JS) estimation method is proposed by Charles Stein in 1956 and achieves a lower estimation error than the MLE method when estimating the true means for a high dimension problem. Thus, the JS method could be better than the MLE method because it shrinks the sample mean towards a better central point.

To better understand how the JS shrinkage works, Efron and Morris [11] use the JS method to estimate the batting abilities of eighteen major-league baseball players. The batting average is calculated as a percentage of hits out of 45 times at the bat. Figure 3.1 shows the observed batting averages shrink toward the grand batting average of 0.265 for each baseball player. It is important to note that the players with extremely low or extremely high observed batting averages experience a higher shrinkage in estimates than the players with batting averages around the grand average. Moreover, the high-performance players expe-

rience a positive shrinkage towards the grand average, while the low-performance players experience a negative shrinkage towards the grand average. Thus, the JS estimator shrinks and adjusts all the estimated values toward a better central point to achieve a better overall estimation than the simple averaging method.

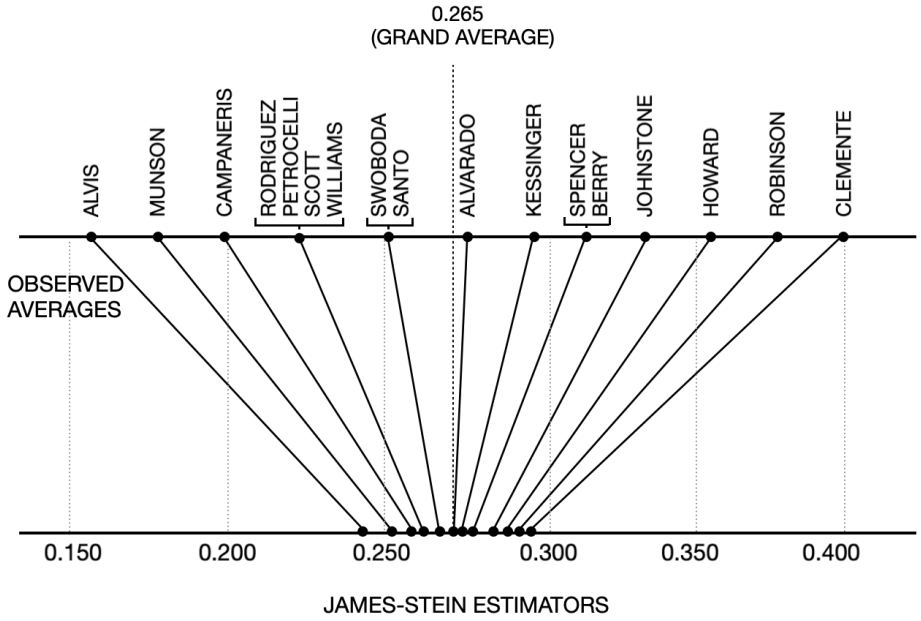


Figure 3.1: James-Stein estimation shrinkage chart, showing all estimates shrink toward the grand average of 0.265 *Note.* Reprinted from p. 121 in Efron and Morris [11].

Furthermore, Efron [8] shows a schematic diagram on how the James-Stein estimator learns from others to update the estimates under a high dimension environment. Figure 3.2 shows that if there are N population means need to be estimated, then the "other" normal statistics $(\hat{M}, \hat{A})$ produced by $(N-1)$ sample means can be used as prior distributions for Case 1 mean estimation. Thus, the Case 1 James-Stein estimator learns from other samples in the data set and itself to create the estimate for Case 1. In conclusion, the JS estimation method takes a portion of the other sample estimates in the set and a portion of its own observed average to shrink the observed average towards the global mean. This concept begins to sound like the inspiration from credibility theory introduced in Chapter 2, such that the credibility premium equals to a portion of the historical loss value and the remaining portion of the observed loss value. The next few subsections explain the JS estimator through more detailed mathematical expressions and derivations, and the similarities between the credibility theory and the JS estimator will be discussed later again.

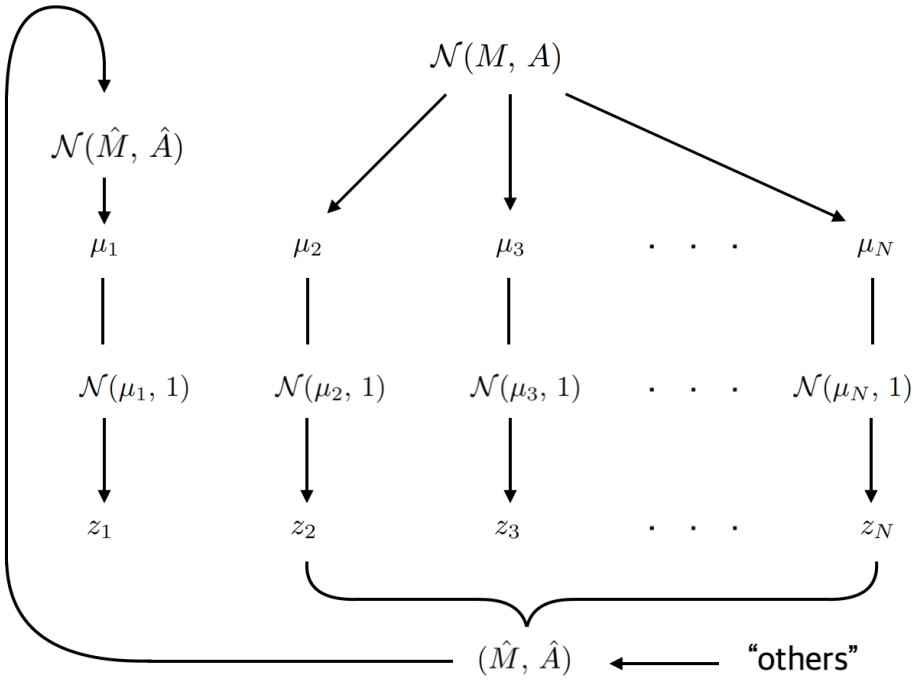


Figure 3.2: Diagram of James-Stein estimation, showing that Case 1 learns from the other $(N-1)$ cases to estimate the mean of Case 1 *Note*. Reprinted from p. 11 in Efron [8].

3.1.2 Stein's Paradox

Stein [19] has proved that the maximum likelihood estimator (MLE) for the mean of a p-dimension random vector with all p entries being independent of each other is admissible for $p = 1, 2$ and inadmissible for $p \geq 3$. Specifically, an admissible estimator is the best possible estimator that has the lowest squared error loss among all the other estimators, while an inadmissible estimator is not the best estimator and there exists some other estimator with a lower squared error loss than it. Thus, Stein [19] has proved that there exist better estimators than the MLE when the estimated random vector has three or more dimensions. This finding is astonishing, because the MLE for the mean of p-dimension random vector simply takes the mean of each entry in the random vector independently, and this naive method has been working well in the past studies. Thus, Stein's paradox explores the fact that when three or more independently distributed entries in a random vector are estimated simultaneously, there is an estimator that is better than the naive method. Furthermore, James [14] has joined Stein's exploration and both find an inadmissible estimator in the explicit form that has a lower expected risk than the MLE when $p \geq 3$.

Let the MLE be $\hat{\boldsymbol{\mu}}^{(\mathrm{MLE})} = \boldsymbol{z}$, where $\boldsymbol{z}$ is a p-dimension normal random vector with all entries being independent of each other, such that $\boldsymbol{z} \sim \mathcal{N}_p(\boldsymbol{\mu},\, I)$ and I is a $p \times p$ identity

matrix. The famous James-Stein estimator is

$$\hat{\boldsymbol{\mu}}^{(\mathrm{JS})} = \left(1 - \frac{p-2}{\|\boldsymbol{z}\|^2}\right) \cdot \boldsymbol{z}, \tag{3.1}$$

where $(1-(p-2)/\|\boldsymbol{z}\|^2)$ is a shrinkage factor, and $\|\boldsymbol{z}\|^2 = \boldsymbol{z}'\boldsymbol{z}$ measures the magnitude of $\boldsymbol{z}$. If the magnitude $\|\boldsymbol{z}\|^2$ is large, then the shrinkage factor is close to 1, and the James-Stein estimator $\hat{\boldsymbol{\mu}}^{(\mathrm{JS})}$ is almost indifferent to the MLE $\boldsymbol{z} = \hat{\boldsymbol{\mu}}^{(\mathrm{MLE})}$. However, if the magnitude $\|\boldsymbol{z}\|^2$ is small then the shrinkage factor is close to 0 and the shrinkage of $\boldsymbol{z}$ towards the origin is large. The reason to shrink $\boldsymbol{z}$ is that samples from a high dimensional random vector are likely to be more scattered away from the origin compared to the true mean of the distribution, and the shrinkage of $\boldsymbol{z}$ towards the origin can decrease the expected squared error loss [14]. Therefore, the James-Stein estimator works better with small-magnitude sample values, and standardization is important in the data step to ensure appropriate shrinkage.

The method of shrinking means towards the origin is the central idea of James and Stein [14], which also introduces a bias to the James-Stein estimator. However, a little sacrifice in bias can exchange an improvement in variance reduction according to the bias-variance trade-off theory. Thus, the biased James-Stein estimator has lower total squared error than the maximum likelihood estimator due to the reduced variance benefiting from the shrinkage, such that

$$E(\|\hat{\boldsymbol{\mu}}^{(\mathrm{JS})} - \boldsymbol{\mu}\|^2) < E(\|\hat{\boldsymbol{\mu}}^{(\mathrm{MLE})} - \boldsymbol{\mu}\|^2).$$

3.1.3 Derivation of the Original James-Stein Estimator by Empirical Bayes

The empirical Bayes estimator can be used to visualize the formation of the original James-Stein estimator [8]. Let μ_i be a normal random variable with mean 0 and variance A, and z_i given μ_i be a normal random variable with mean μ_i and variance 1 for $i = 1, \ldots, p$. Specifically, $\mu_i \sim \mathcal{N}(0,\, A)$ and $z_i|\mu_i \sim \mathcal{N}(\mu_i,\, 1)$. Then it can be shown easily by the Bayes theorem that the posterior distribution of μ_i given z_i is normally distributed with mean $B \cdot z_i$ and variance B, that is, $\mu_i|z_i \sim \mathcal{N}(B \cdot z_i,\, B)$, where $B = A/(A+1)$. The marginal distribution of z_i can be obtained by integrating the joint density of z_i and μ_i (the product of the density of μ_i and the density of $z_i|\mu_i$) with respect to μ_i to get $z_i \sim \mathcal{N}(0,\, A+1)$.

Now let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ be a p-dimension random vector that is normally distributed with mean vector $\mathbf{0}$ and covariance matrix $A \cdot I$, such that $\boldsymbol{\mu} \sim \mathcal{N}_p(\mathbf{0}, A \cdot I)$, where I is a $p \times p$ identity matrix. Next, let $\boldsymbol{z} = (z_1, \ldots, z_p)'$ given $\boldsymbol{\mu}$ be p-dimension normally distributed, such that $\boldsymbol{z}|\boldsymbol{\mu} \sim \mathcal{N}_p(\boldsymbol{\mu},\, I)$, where the $(\mu_i,\, z_i)$ pairs for $i = 1, \ldots, p$ are independent of each other. By the Bayes theorem, the posterior distribution of $\boldsymbol{\mu}$ given $\boldsymbol{z}$ can be expressed as

follows:

$$\boldsymbol{\mu}|\boldsymbol{z} \sim \mathcal{N}_p(B \cdot \boldsymbol{z}, B \cdot I),$$

where $B \cdot \boldsymbol{z}$ is the Bayes estimator, denoted by $\hat{\boldsymbol{\mu}}^{(\text{Bayes})}$, for a p-dimension random vector with unknown A from $\boldsymbol{\mu}$'s covariance matrix $A \cdot I$, that is,

$$\hat{\boldsymbol{\mu}}^{(\text{Bayes})} = B \cdot \boldsymbol{z} = \frac{A}{A+1} \cdot \boldsymbol{z} = \left(1 - \frac{1}{A+1}\right) \cdot \boldsymbol{z}.$$

Next, we will show that the unknown fraction $1/(A+1)$ in $\hat{\boldsymbol{\mu}}^{(\text{Bayes})}$ can be replaced with $(p-2)/\|\boldsymbol{z}\|^2$. First, the marginal distribution of $\boldsymbol{z}$, $\boldsymbol{z} \sim \mathcal{N}_p(0,\, (A+1) \cdot I)$, which implies $\boldsymbol{z}/\sqrt{A+1} \sim \mathcal{N}_p(0,\, I)$ and

$$\|\boldsymbol{z}\|^2/(A+1) = \|\boldsymbol{z}/\sqrt{A+1}\|^2 = (z_1^2 + z_2^2 + \cdots + z_p^2)/(A+1) \sim \chi_p^2.$$

Second, it is known that the expectation of the inverse of χ_p^2 (a chi-squared random variable with p degrees of freedom) is the inverse of $p-2$, such that

$$E\left[\frac{1}{(A+1) \cdot \chi_p^2}\right] = \frac{1}{(A+1) \cdot (p-2)}.$$

Finally, the expected value of $(p-2)/\|\boldsymbol{z}\|^2$ can be expressed as

$$E\left[\frac{p-2}{\|\boldsymbol{z}\|^2}\right] = \frac{p-2}{(A+1) \cdot (p-2)} = \frac{1}{A+1},$$

which explains the replacement of $1/(A+1)$ in the Bayes estimator $\hat{\boldsymbol{\mu}}^{(\text{Bayes})}$ by $(p-2)/\|\boldsymbol{z}\|^2$. Therefore, the original James-Stein estimator can be re-expressed as

$$\hat{\boldsymbol{\mu}}^{(\text{JS})} = \left(1 - \frac{p-2}{\|\boldsymbol{z}\|^2}\right) \cdot \boldsymbol{z} = \left(1 - \frac{1}{A+1}\right) \cdot \boldsymbol{z} = B \cdot \boldsymbol{z} = \hat{\boldsymbol{\mu}}^{(\text{Bayes})}.$$

3.1.4 Estimate the Mean of a Multivariate Normal Random Vector

Since the James-Stein estimator leads to a lower squared error loss than the MLE when the dimension of the underlying multivariate normal random vector is equal or more than 3, it is a good estimator to estimate the means of many independently distributed multivariate random vectors. Specifically, modeling of mortality rates involves many independent random vectors for a wide range of study ages and years. Therefore, the JS method needs to be modified under a multivariate setting. Fortunately, Jorion [15] provides an alternate form of the James-Stein estimator that can be applied directly to multivariate random vectors.

Let $\underline{Y}_t = (Y_{1,t}, \ldots, Y_{p,t})'$, $t = 1, \ldots, T$, be a sample of T random vectors from a p-dimension normal distribution with unknown mean $\underline{\boldsymbol{\mu}}$ and known covariance matrix $\boldsymbol{\Sigma}$. The problem is to estimate the unknown mean $\underline{\boldsymbol{\mu}}$. Let $\underline{\mathbf{y}} = (\underline{y}_1, \ldots, \underline{y}_T)$ be a realization of the

random sample $\underline{\mathbf{Y}} = (\underline{Y}_1, \ldots, \underline{Y}_T)$. When p is equal or larger than 3, Efron and Morris [10] show that the maximum likelihood estimator $\underline{\hat{\boldsymbol{\mu}}}^{MLE}(\underline{\mathbf{y}}) = \underline{\bar{\mathbf{y}}}$, also the vector of sample means (specifically, $\underline{\bar{\mathbf{y}}} = \underline{\hat{\boldsymbol{\mu}}}^{MLE}(\underline{\mathbf{y}}) = (1/T) \cdot \sum_{t=1}^{T} \underline{y}_t$), is inadmissible subject to the following quadratic loss function

$$L[\underline{\hat{\boldsymbol{\mu}}}, \underline{\hat{\boldsymbol{\mu}}}(\underline{\mathbf{y}})] = [\underline{\hat{\boldsymbol{\mu}}} - \underline{\hat{\boldsymbol{\mu}}}(\underline{\mathbf{y}})]' \cdot \Sigma^{-1} \cdot [\underline{\hat{\boldsymbol{\mu}}} - \underline{\hat{\boldsymbol{\mu}}}(\underline{\mathbf{y}})]. \tag{3.2}$$

Then, the James-Stein estimator $\underline{\hat{\boldsymbol{\mu}}}^{JS}(\underline{\mathbf{y}})$ for this setting, called a modified James-Stein estimator, can be expressed as

$$\underline{\hat{\boldsymbol{\mu}}}^{JS}(\underline{\mathbf{y}}) = (1 - \hat{w}) \cdot \underline{\bar{\mathbf{y}}} + \hat{w} \cdot y_0 \cdot \underline{\mathbf{1}}, \tag{3.3}$$

where $\underline{\mathbf{1}}$ is a $(p \times 1)$ unit vector; y_0 is the overall mean over all entries in the matrix of the entire random sample $\underline{\mathbf{y}}$, that is,

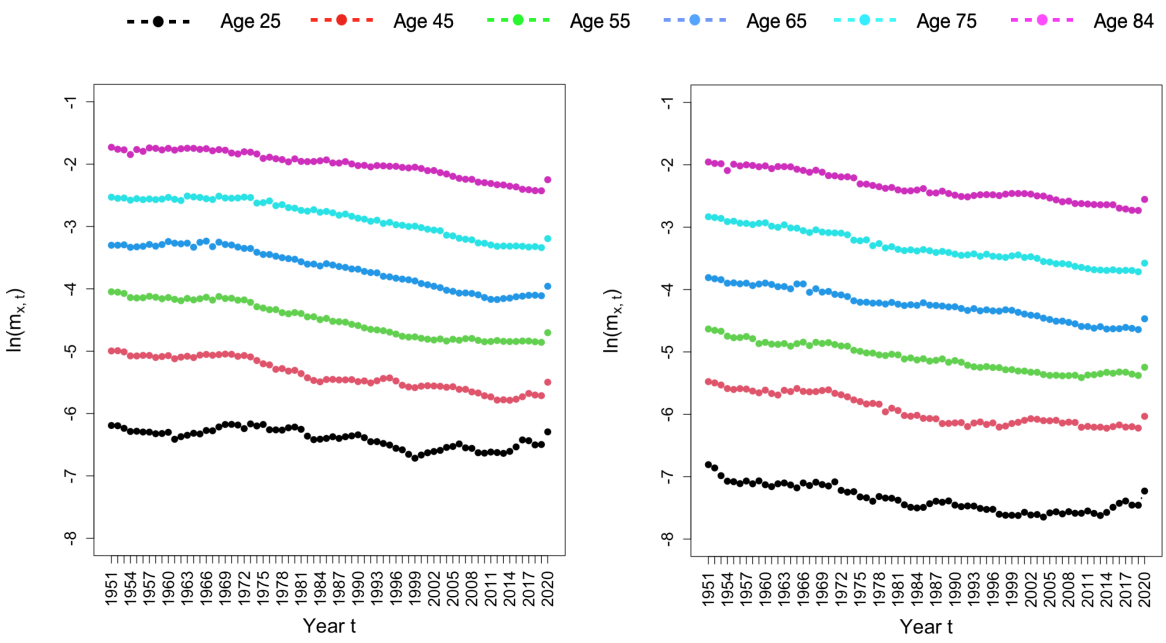
$$y_0 = \frac{1}{p \cdot T} \cdot \sum_{n=1}^{p} \sum_{t=1}^{T} y_{n,t};$$

and

$$\hat{w} = \min \left[1, \frac{(p-2)/T}{(\underline{\bar{\mathbf{y}}} - y_0 \cdot \underline{\mathbf{1}})' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\underline{\bar{\mathbf{y}}} - y_0 \cdot \underline{\mathbf{1}})} \right].$$

Note that firstly, the modified James-Stein estimator (also called a shrinkage estimator) with the explicit form (3.3) has an extra term compared to the expression (3.1) for the original James-Stein estimator. Secondly, the explicit form (3.3) equals a weighted average of the vector of sample means ($\underline{\bar{\mathbf{y}}}$) and the vector of overall mean with a shrinkage factor $(1 - \hat{w})$, which takes a value between 0 and 1. Thus, the modified James-Stein estimator equals the vector of overall mean (the vector of sample means), as an estimate of the unknown mean $\underline{\boldsymbol{\mu}}$, when $\hat{w} = 1$ ($\hat{w} = 0$). Thirdly, if $\underline{\bar{\mathbf{y}}}$ is closer to $y_0 \cdot \underline{\mathbf{1}}$ (that is, the sample means in all entries of $\underline{\bar{\mathbf{y}}}$ are closer to y_0), which implies $L[\underline{\bar{\mathbf{y}}}, y_0 \cdot \underline{\mathbf{1}}]$ (the deviation measured by the quadratic loss function between $\underline{\bar{\mathbf{y}}}$ and $y_0 \cdot \underline{\mathbf{1}}$ in (3.2)) is smaller, then $\hat{w}$ is larger and a more weight is placed on $y_0 \cdot \underline{\mathbf{1}}$ (the vector of overall mean). On the contrary, if $\underline{\bar{\mathbf{y}}}$ is far away from $y_0 \cdot \underline{\mathbf{1}}$ (that is, the sample means in some entries of $\underline{\bar{\mathbf{y}}}$ are far away from y_0, or the deviation is big), then $\hat{w}$ is small and a large weight is placed on $\underline{\bar{\mathbf{y}}}$ (the vector of sample means). Finally, if T is large (small), that is, more (less) $\underline{y}_t$ is sampled, then $\hat{w}$ is small (large) and a heavy (light) weight is placed on $\underline{\bar{\mathbf{y}}}$, which is consistent with the intuition that more (less) sampled data give a more (less) reliable/accurate estimate for the unknown true mean $\underline{\boldsymbol{\mu}}$.

It is also interesting to note that the expression for the modified JS estimator $\hat{\boldsymbol{\mu}}^{(\mathrm{JS})}$ is almost identical to the credibility theory formula for the risk premium (RP) introduced in



(a) U.S. Males (b) U.S. Females

Figure 3.3: Downward time trends of $\ln(m_{x,t})$ from $t = 1951$ to $t = 2020$ for six ages

Section 2.4 as

$$RP = Z \cdot D + (1 - Z) \cdot M,$$

where the historical manual loss value M is replaced by y_0 (the overall mean of the sample data); the observed loss value D is replaced by a vector of means, over all $t = 1, \ldots, T$ in the random sample, for each entry; and the credibility factor Z is replaced by the shrinkage factor $(1 - \hat{w})$. Therefore, adopting the modified James-Stein estimator is inspired from the credibility theory both conceptually and mathematically.

3.2 Application of the James-Stein Estimator

In order to apply the modified James-Stein estimator to the mortality data, it is important to understand the underlying random variables. In this project, the mortality rate for each age x is considered a random variable and the mortality rate for that age is measured repeatedly throughout the entire fitting years. The ultimate goal is to predict the mortality rates for each age over the forecasting years.

The famous Lee-Carter model adopts $\ln(m_{x,t})$ (the natural logarithm of central rate for age x in year t) to predict mortality rates. Historical data from the Human Mortality Database (www.mortality.org) show that there is a downward time trend in $\ln(m_{x,t})$, which makes the mortality data $\ln(m_{x,t})$ non-stationary, and it is essential to eliminate the downward trend before forecasting using a decrement processing. For instance, Figure 3.3

exhibits that $\ln(m_{x,t})$s for both genders of the U.S. for six selected ages (25, 45, 55, 65, 75, 84) from 1951 to 2020 display a decreasing trend in the past 70 years except a slight increase in 2020 largely due to Covid-19 pandemic.

In this project, the mortality data $m_{x,t}$s are obtained from the Human Mortality Database. In order to apply the modified James-Stein estimator to the mortality rate forecasting problem, the first step is to eliminate the downward time trend of $\ln(m_{x,t})$s, we take the first-order difference on $\ln(m_{x,t})$s with respect to t, that is,

$$Y_{x,t} = \ln(m_{x,t+1}) - \ln(m_{x,t}), \quad x = x_L, \ldots, x_U, \quad t = t_L, \ldots, t_U - 1,$$

where $[x_L,\, x_U] \times [t_L,\, t_U]$ is the fitting age-year window. Note that $Y_{x,t}$ can be interpreted as the mortality decrement (mortality improvement) for age x in year t. To apply the modified James-Stein estimator (3.3), let

$$\underline{\mathbf{Y}} = (\underline{Y}_1, \ldots, \underline{Y}_T) = \begin{bmatrix} Y_{x_L,t_L} & \ldots & Y_{x_L,t_U-1} \\ \vdots & \ddots & \vdots \\ Y_{x_U,t_L} & \ldots & Y_{x_U,t_U-1} \end{bmatrix}_{p \times T},$$

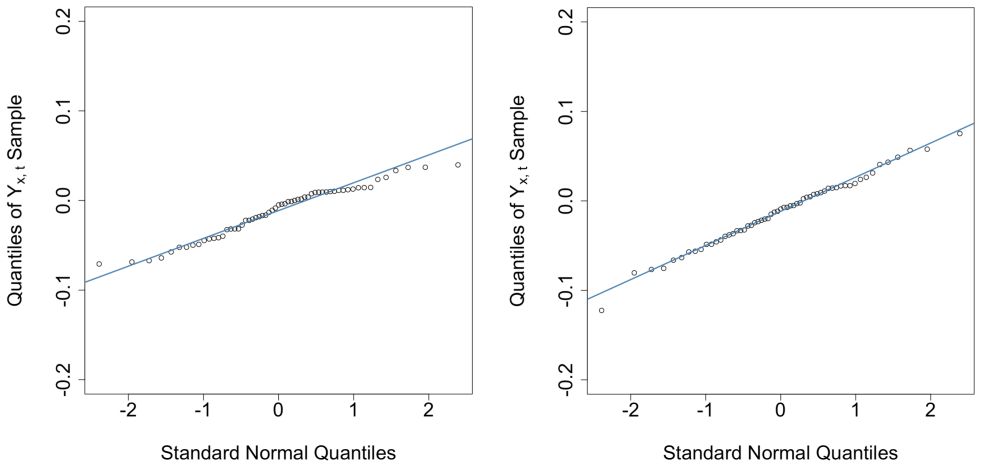
where $p = x_U - x_L + 1$ and $T = t_U - t_L$. To satisfy the normal assumption for the modified James-Stein estimator, that is, $Y_{x,t}$, $t = t_L, \ldots, t_U - 1$, are assumed to be identical and independent normal random variables with unknown mean μ_x to be estimated for $x = x_L, \ldots, x_U$. It is essential to check the normality of $Y_{x,t}$, $t = t_L, \ldots, t_U - 1$. Figure 3.4 displays the Q-Q plots of $Y_{x,t}, t = t_L, \ldots, t_U - 1$ for $x = 45$ years old against the standard normal quantiles for all six populations. The Q-Q plots show that the $Y_{x,t}$, $t = t_L, \ldots, t_U - 1$ quantiles form straight lines when plot against the standard normal quantiles for all six populations. Therefore, it is reasonable to assume the random variables $Y_{x,t}$, $t = t_L + 1, \cdots, t_U$ are normally distributed.

Furthermore, the variance $\hat{\sigma}_x^2$ and the covariance $\hat{\sigma}_{x_i,x_j}$ $(x_i \neq x_j)$ in the $p \times p$ variance-covariance matrix of $Y_{x_L,t}, \ldots, Y_{x_U,t}$,

$$\mathbf{\Sigma} = \begin{bmatrix} \hat{\sigma}_{x_L}^2 & \ldots & \hat{\sigma}_{x_L,x_U} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{x_L,x_U} & \ldots & \hat{\sigma}_{x_U}^2 \end{bmatrix},$$

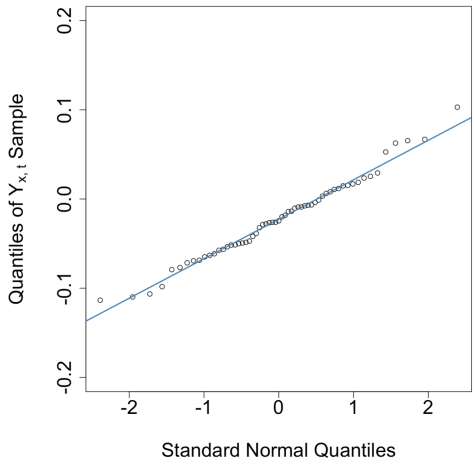
can be estimated by

$$\hat{\sigma}_x^2 = \frac{1}{T-1} \sum_{t=t_L}^{t_U} (Y_{x,t} - \bar{Y}_x)^2, \quad x = x_L, \ldots, x_U,$$

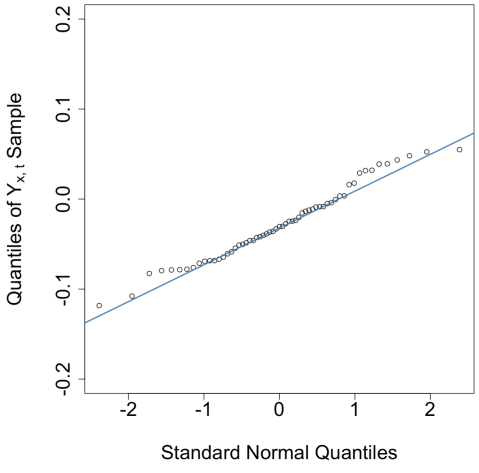


(a) US Male

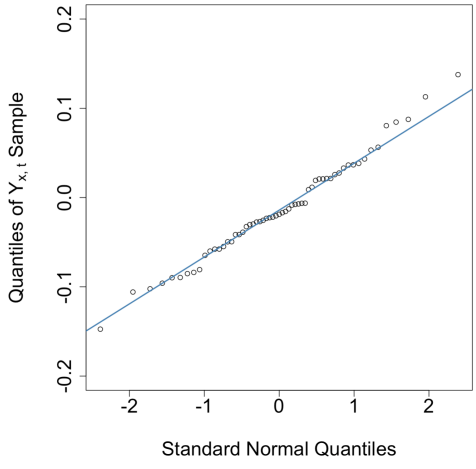
(b) US Female



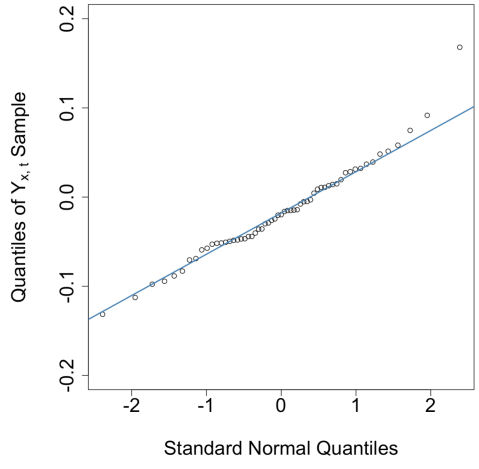
(c) Japan Male



(d) Japan Female



(e) UK Male



(f) UK Female

Figure 3.4: Q-Q plots for $Y_{x,t}$ with $x = 45$

and

$$\hat{\sigma}_{x_i,x_j} = \frac{1}{T-1}\sum_{t=t_L}^{t_U}(Y_{x_i,t} - \bar{Y}_{x_i}) \cdot (Y_{x_j,t} - \bar{Y}_{x_j}), \quad x_i,\, x_j = x_L, \ldots, x_U, \quad x_i \neq x_j,$$

where $\bar{Y}_x = (1/T)\sum_{t=t_L}^{t_U} Y_{x,t}$.

Note that each row vector $(Y_{x,t_L}, \ldots, Y_{x,t_U-1})$ in $\underline{\mathbf{Y}}$ is an individual sequence of yearly mortality decrements for age x, and $(1/p)\sum_{x=x_L}^{x_U}(Y_{x,t_L}, \ldots, Y_{x,t_U-1})$ is the group sequence of average yearly mortality decrements for all ages. In the context of mortality rates, the vector of the sample means, $\bar{\underline{\mathbf{Y}}} = \hat{\underline{\boldsymbol{\mu}}}^{MLE}(\underline{\mathbf{Y}}) = (1/T)\cdot\sum_{t=1}^{T}\underline{Y}_t$, in (3.3) is the vector of the averages of annual mortality decrements; and Y_0, in the vector of the overall mean, $Y_0 \cdot \underline{\mathbf{1}}$, which takes the average over all entries of $\underline{\mathbf{Y}}$ is the overall annual mortality decrement for the group sequence. Therefore, the modified James-Stein estimator in (3.3) gives a weighted average of the average annual mortality decrement for each age-specific individual time sequence and the average of annual mortality decrement for the group time sequence. If $\hat{w}$ is larger (smaller) than 0.5, implied by that T is smaller (bigger) or the sample means for all ages are more (less) centered around the overall mean), then a heavier (lighter) weight is placed on the overall mean Y_0; that is, the group time sequence contributes its average annual mortality decrement to the estimated annual mortality decrement more (less) than each age-specific individual time sequence.

Once the annual mortality decrement for age x is estimated from the modified James-Stein estimator (3.3), the natural logarithm of central death rate for year $t_U + n$ (n years away from the last fitting year t_U) can be predicted as

$$\ln(\hat{m}_{x,t_U+n}) = \ln(m_{x,t_U}) + n \times \hat{Y}_x, \quad n = 1, 2, \ldots, T_0$$

where $\hat{Y}_x$ is the entry for age x in the vector of the modified James-Stein estimator $\hat{\underline{\boldsymbol{\mu}}}^{JS}(\underline{\mathbf{Y}}) = (1-\hat{w})\cdot\bar{\underline{\mathbf{Y}}} + \hat{w}\cdot Y_0\cdot\underline{\mathbf{1}}$, and T_0 is the number of forecasting years.

Finally, the natural logarithm of the predicted central death rate for year $t_U + n$, $\ln(\hat{m}_{x,t_U+n})$, is converted back to the predicted one-year death probability by first taking the exponential on $\ln(\hat{m}_{x,t_U+n})$ to get $\hat{m}_{x,t_U+n}$ and then

$$\hat{q}_{x,t_U+n} = 1 - \exp(-\hat{m}_{x,t_U+n}).$$

The equality above is based on the assumption of the constant force of mortality (CF) within a year, that is, $\mu_{x,t} = \mu_{x+s,t+s}$ for $s \in [0,1)$, which is interpreted as the force of mortality $\mu_{x,t}$ is constant within a year for each integer age x and integer year t. Then

$$q_{x,t} = 1 - p_{x,t} = 1 - \exp(-\int_0^1 \mu_{x+s,t+s}\, ds) = 1 - \exp(-\mu_{x,t}),$$

and

$$m_{x,t} = \frac{q_{x,\,t}}{\int_0^1 {}_s p_{x,\,t}\, ds} = \frac{q_{x,\,t}}{\int_0^1 e^{-\int_0^s \mu_{x+r,\,t+r}\, dr} ds} = \frac{q_{x,\,t}}{\int_0^1 e^{-s\,\mu_{x,\,t}} ds} = \frac{q_{x,\,t}}{(1 - e^{-\mu_{x,\,t}})/\mu_{x,\,t}} = \mu_{x,\,t}.$$

Chapter 4

Numerical Illustrations

The James-Stein estimation method is applied to the mortality data set for both genders of the U.S., the U.K., and Japan. The goal is to use the modified JS estimator to forecast mortality rates for 10, 20, and 30 years. The mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE) are adopted to evaluate the forecasting performance of the modified JS estimator against the other underlying mortality models which are Lee-Carter (LC) model, Cairns-Blake-Dowd (CBD) model, M6 and M7 models (two extensions of CBD model), and Renshaw-Haberman (RH) model.

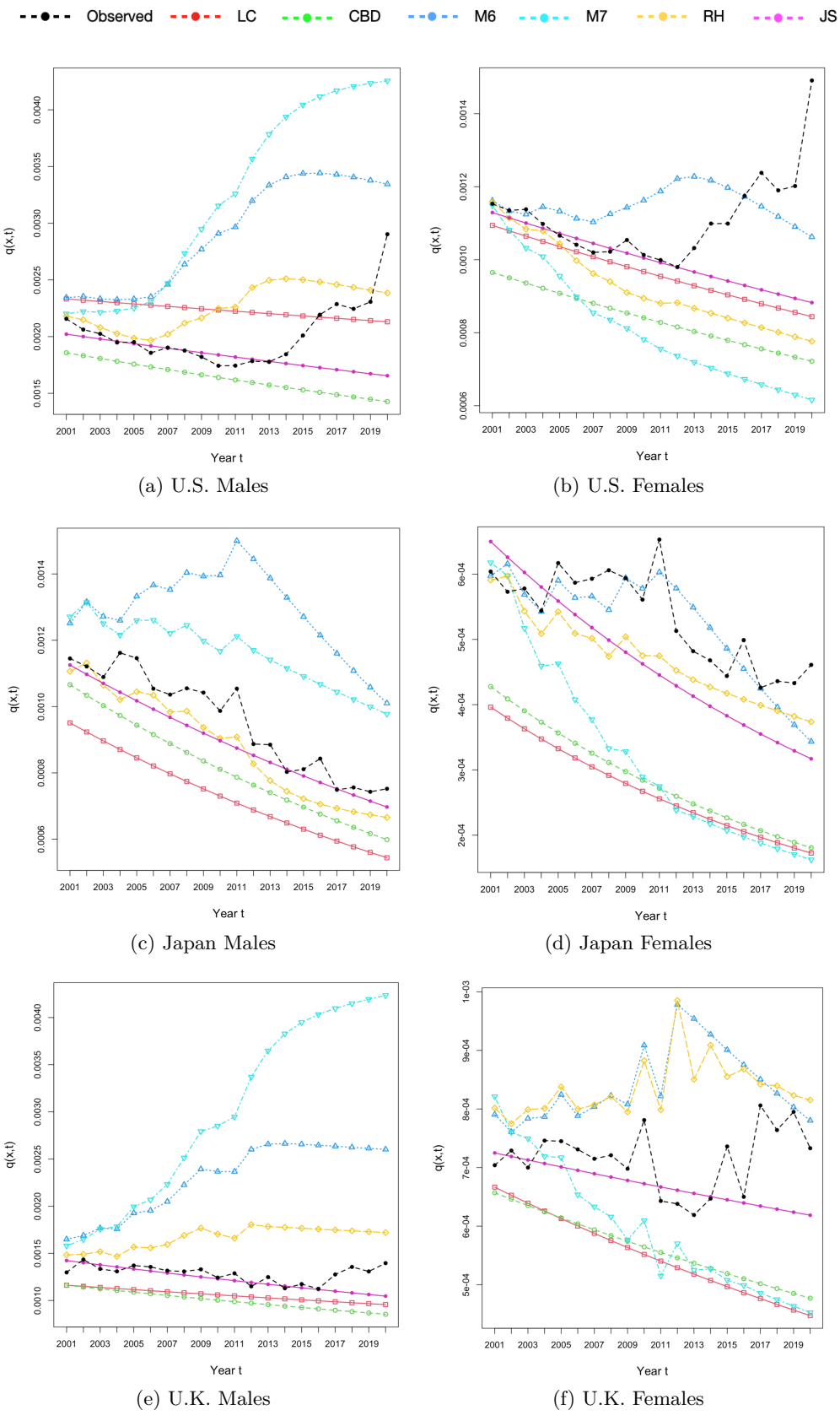
The mortality data used in this project is extracted from the Human Mortality Database (HMD, www.mortality.org) with ages ranging from 25 to 84 and years ranging from 1951 to 2020. The training and test data sets are given in Table 4.1. Thus, there are in total 108 combinations to be implemented for two genders, three countries, three lengths of forecasting years, and six mortality models.

Forecast Years T_0	Training Data Set $[x_L,\, x_U] \times [t_L,\, t_U]$	Test Data Set $[x_L,\, x_U] \times [t_U + 1,\, t_U + T_0]$
10 years	$[25, 84] \times [1951, 2010]$	$[25, 84] \times [2011, 2020]$
20 years	$[25, 84] \times [1951, 2000]$	$[25, 84] \times [2001, 2020]$
30 years	$[25, 84] \times [1951, 1990]$	$[25, 84] \times [1991, 2020]$

Table 4.1: Training and test data sets used for mortality modeling

4.1 Prediction of Mortality Rates

The observed and predicted one-year mortality rates $q_{x,\,t}$ using six different mortality models are illustrated in this section. First, a young age 37 and an old age 71 are selected to show the differences in the predicted mortality rate $q_{x,\,t}$ among six mortality models. Figures 4.1 and 4.2 display the observed and predicted mortality rate $q_{x,\,t}$ curves for the forecast year t from 2001 to 2020 for ages 37 and 71 respectively. Figure 4.3 also shows the average observed and predicted mortality rate $\bar{q}_t$ $(= (1/p) \sum_{x=x_L}^{x_U} q_{x,\,t})$ curves for $t = 2001, \ldots, 2020$. The observations are summarized below:



(a) U.S. Males

(b) U.S. Females

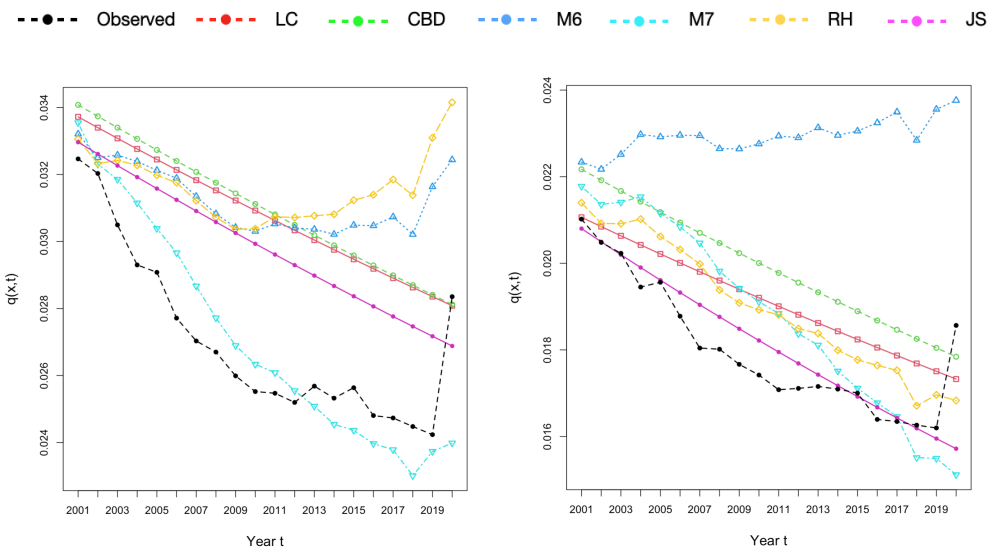
(c) Japan Males

(d) Japan Females

(e) U.K. Males

(f) U.K. Females

Figure 4.1: $q_{37,t}$ against $t = 2001, \ldots, 2020$

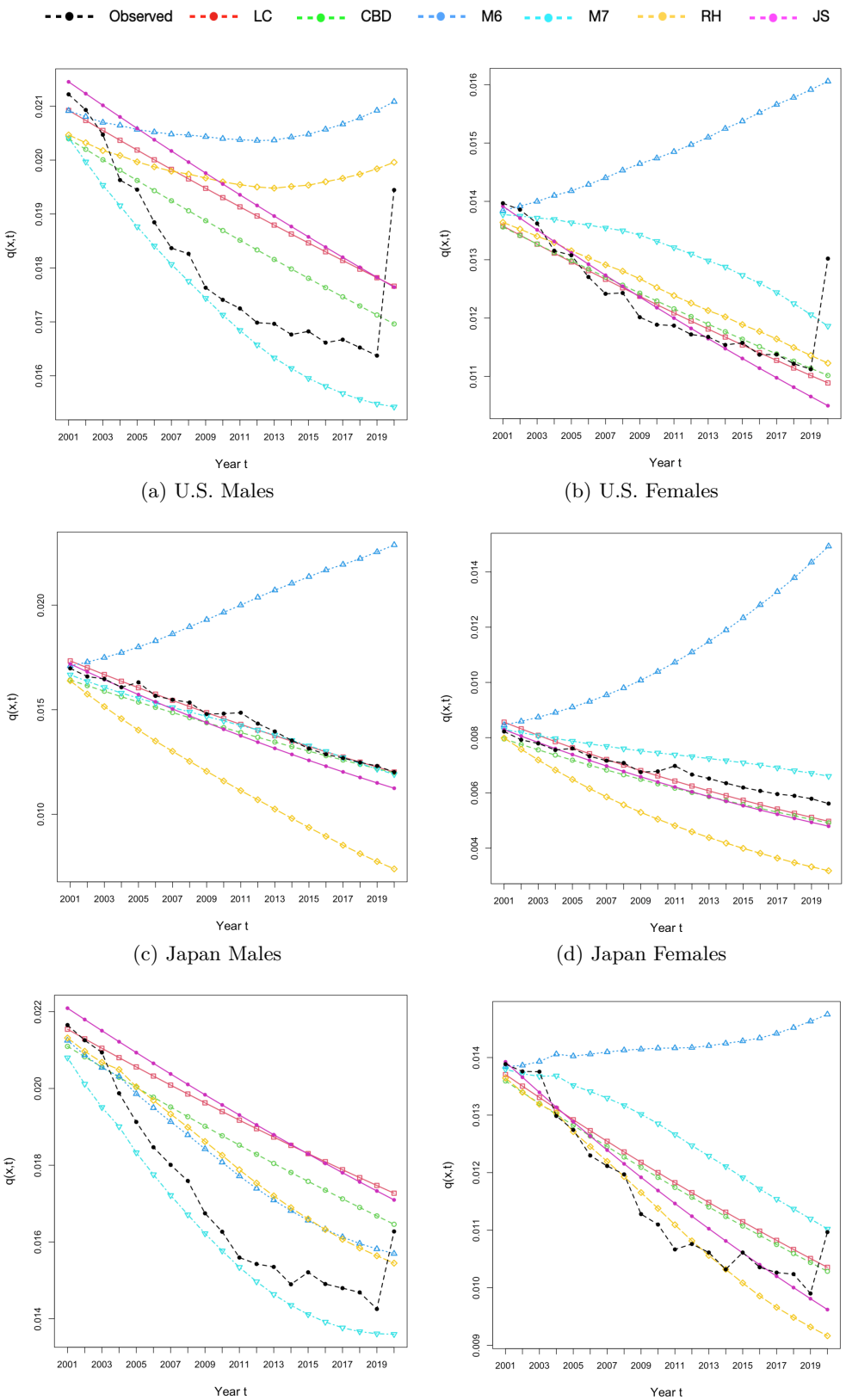


(a) U.S. Males (b) U.S. Females

(c) Japan Males (d) Japan Females

(e) U.K. Males (f) U.K. Females

Figure 4.2: $q_{71,t}$ against $t = 2001, \ldots, 2020$



(a) U.S. Males

(b) U.S. Females

(c) Japan Males

(d) Japan Females

(e) U.K. Males

(f) U.K. Females

Figure 4.3: $\bar{q}_t = (1/60)\sum_{x=25}^{84} q_{x,t}$ against $t = 2001, \ldots, 2020$

- The 37 years old individuals have lower observed and predicted mortality rates than the 71 years old individuals in Figures 4.1 and 4.2. Moreover, the 37 years old U.S. individuals experience deteriorating mortality rates since 2014 and on wards, while the 71 years old U.S. individuals have steady mortality rate improvement from 2001 to 2020. Next, the 37 years old Japan individuals have improved mortality rates despite the minor fluctuations in mortality rates from 2001 to 2020. Finally, both genders of the 37 years old U.K. individuals experience flat mortality rates from 2001 to 2020 without much mortality rate improvement or deterioration, while both genders of the 71 years old U.K. individuals continuously get mortality rate improvement as time passes by.

- The LC, CBD, and JS mortality models result in smooth downward prediction trends, while the M6, M7, and RH models produce fluctuating and random mortality rate predictions. Therefore, the former three models perform better when the observed mortality rates are relatively stable, and the latter three models work better when the observed mortality rates fluctuate drastically across years.

- The JS model predicts $q_{37,t}$s quite well for both genders of the 37 years old U.S. and U.K. individuals compared to the other models, especially the pink JS prediction line for U.K. males is very close to the black line for the observed values. Furthermore, it is hard to conclude the best forecasting model for Japan male and female populations from Figure 4.1, and we need more performance evaluation metrics, which are introduced in the next section.

- The JS model predicts $q_{71,t}$s very well for both genders of the 71 years old Japan individuals and the 71 years old U.S. females compared to the other models. On the other hand, the M7 model in cyan color, with an age-year cohort effect and especially designed for older ages, predicts $q_{71,t}$s quite well for U.S. males and U.K. males. Finally, the RH model in yellow color, which has an age-year cohort effect, predicts better $q_{71,t}$s for U.K. females than the other models.

- From Figure 4.3, we observe that the average observed $\bar{q}_t$ curves have a sharp jump in 2020 for U.S. and U.K. It is interesting to observe that none of six mortality models can catch these sudden mortality rate jumps, largely due to the Covid-19 pandemic crisis.

- The average observed $\bar{q}_t$ curves across six populations in Figure 4.3 look similar to those for age 71 in Figure 4.2. This is because $q_{x,t}$s increase steadily as age x increases, and the older individuals have much higher $q_{x,t}$s than the younger individuals. As a result, the $q_{x,t}$s for old individuals are closer to $\bar{q}_t$. Figure 4.3 for $\bar{q}_t$ versus year t shows that the James-Stein model works well for U.S. females and both genders of Japan on average.

Therefore, the plots of mortality rate $q_{x,t}$ against year t give a general idea of the predictive performances of models; the James-Stein model predicts $q_{x,t}$s for young U.S. and U.K. populations better than the other models. However, more performance metrics calculations are needed to quantitatively conclude the performance of each mortality model at a wide range of ages for different populations.

4.2 Performance Metrics

Three performance metrics are used to summarize the differences between the observed and forecast mortality rates across all study ages $[x_L, x_U]$ and forecasting years $[t_U+1, t_U+T_0]$. They are the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE), which are defined as follows:

- Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{T_0 \cdot (x_U - x_L + 1)} \sum_{n=1}^{T_0} \sum_{x=x_L}^{x_U} |\hat{q}_{x,\,t_U+n} - q_{x,\,t_U+n}|,$$

- Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{T_0 \cdot (x_U - x_L + 1)} \sum_{n=1}^{T_0} \sum_{x=x_L}^{x_U} \left| \frac{\hat{q}_{x,\,t_U+n} - q_{x,\,t_U+n}}{q_{x,\,t_U+n}} \right|,$$

- Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{T_0 \cdot (x_U - x_L + 1)} \sum_{n=1}^{T_0} \sum_{x=x_L}^{x_U} (\hat{q}_{x,\,t_U+n} - q_{x,\,t_U+n})^2},$$

where $\hat{q}_{x,t}$ and $q_{x,t}$ are the predicted and observed mortality rates, respectively, for age x in year t, t_U is the last fitting year, and T_0 is the last forecasting year.

4.3 Forecasting Performances

In this section, the performance metrics outputs are given in numerical form with Tables 4.2– 4.4 and in graphical form with Figures 4.4–4.7. The forecasting errors at different ages and forecasting years are summarized into three performance metrics for each of six populations. The observations from the performance metrics tables can be concluded as follows:

- The James-Stein (JS) mortality model has the lowest average MAE, MAPE and RMSE over all six populations and thus the best performance compared to the other five

	Average	U.S.-M	U.S.-F	Japan-M	Japan-F	U.K.-M	U.K.-F
MAE (10^{-3})							
LC	0.7349	1.1942	0.8202	0.4957	0.3765	0.8820	0.6409
CBD	1.2471	1.8236	1.2219	0.9959	0.8413	1.5323	1.0673
RH	1.0864	**0.7322**	0.7981	1.5799	1.0158	1.2382	1.1543
M6	1.2334	0.7767	0.8045	2.2781	1.6968	1.1440	0.7002
M7	1.1700	2.0100	0.8812	**0.4896**	0.7338	2.4876	**0.4179**
JS	**0.5818**	0.9639	**0.5926**	0.5394	**0.3450**	**0.5753**	0.4747
MAPE (%)							
LC	13.26	11.07	13.19	9.28	22.16	11.00	12.85
CBD	16.39	15.27	16.88	11.32	24.01	15.21	15.64
RH	9.01	**5.63**	8.99	13.09	11.34	6.92	8.11
M6	13.34	7.28	**6.40**	23.53	22.80	12.06	7.98
M7	13.95	9.78	15.83	6.81	20.12	18.87	12.30
JS	**7.62**	10.01	9.69	**5.43**	**7.54**	**5.92**	**7.15**
RMSE (10^{-3})							
LC	1.3571	2.0200	1.7076	**0.9051**	**0.4647**	1.5093	1.5359
CBD	2.9863	4.2048	2.8468	2.5211	1.9387	3.7461	2.6602
RH	2.5388	1.8001	1.8293	3.2503	2.4666	3.1102	2.7763
M6	2.5363	**1.6717**	1.4923	4.3459	3.4064	3.0364	1.2652
M7	2.7052	5.0567	**1.3620**	1.1362	1.5045	6.4206	**0.7510**
JS	**1.2417**	1.9009	1.3877	1.0116	0.6063	**1.4260**	1.1175

Table 4.2: Forecasting performance metrics for forecasting 10 years

models when forecasting 10, 20, and 30 years. It's interesting to notice that the JS model is not always the best model for all six populations, but its performance metrics outputs are always at lower levels for each population compared to the other models, which allow the JS model to achieve the overall best forecasting performance.

- The JS model outperforms the other forecasting models across all three performance metrics when we are forecasting 10 years for U.K. males, 20 years for females of all three countries, and 30 years for U.S. males, Japan females, and U.K. females. However, the RH model dominates the other mortality models when forecasting 30 years for U.S. females. On top of this, there is no other forecasting model achieves the lowest forecasting errors across all three performance metrics for any population. Therefore, the JS model achieves the best forecasting results across all three performance metrics for some selected populations.

	Average	U.S.-M	U.S.-F	Japan-M	Japan-F	U.K.-M	U.K.-F
MAE (10^{-3})							
LC	1.4843	2.1903	0.8306	**0.6608**	0.8500	3.1998	1.1747
CBD	1.6259	**2.1499**	1.2840	1.3149	1.0491	2.6652	1.2924
RH	1.8844	2.4353	0.7525	3.6593	1.9739	**1.6699**	0.8153
M6	3.8704	3.1915	3.2812	6.5085	5.0713	1.8294	3.3403
M7	2.2209	2.6461	2.2495	1.3668	1.3388	3.9065	1.8175
JS	**1.3520**	2.2849	**0.5979**	0.8080	**0.5859**	3.0890	**0.7463**
MAPE (%)							
LC	18.80	13.89	11.72	14.66	35.71	18.24	18.54
CBD	21.40	18.45	19.47	15.11	33.81	23.48	18.10
RH	17.04	18.11	11.50	18.39	20.12	21.28	12.82
M6	33.99	26.69	20.54	40.10	51.18	40.18	25.24
M7	30.04	26.51	23.72	13.50	32.88	66.95	16.65
JS	**12.38**	**12.80**	**10.41**	**7.26**	**18.71**	**15.44**	**9.67**
RMSE (10^{-3})							
LC	2.2086	3.2551	1.3033	**0.9682**	1.1045	5.0002	1.6202
CBD	2.5252	**3.0557**	1.9372	2.6371	1.8480	3.6923	1.9807
RH	3.1994	3.0720	1.1912	6.9660	3.9161	2.4334	1.6176
M6	6.3903	4.2608	5.4217	11.6210	9.4235	**2.2794**	5.3356
M7	3.9190	4.9023	3.7796	2.5912	2.1756	6.8069	3.2583
JS	**2.1621**	3.6148	**1.0902**	1.4324	**0.8123**	4.7526	**1.2702**

Table 4.3: Forecasting performance metrics for forecasting 20 years

- All three metrics have lower (higher) error values when the forecasting year is shorter (longer), which is consistent with intuition. This is because longer forecasting years are more unpredictable compared to shorter forecasting years, and lead to higher forecasting errors.

In conclusion, the best performing forecasting model for a population and a forecasting year varies due to different mortality trends at different ages, but the JS model is the best overall performing model based on the three performance metrics analyses.

Although the numerical results in Tables 4.2–4.4 give summarized forecasting performance results from the six models, Figure 4.4 displays the box plots of MAPEs over all 60 study ages, forecasting years and six populations for six models, and Figures 4.5–4.7 exhibit the $MAPE_t$ curves against forecasting year t for forecasting 10, 20 and 30 years, respective, where $MAPE_t$ is the MAPE for year t and calculated only across all study ages. The following are the observations from these figures:

	Average	U.S.-M	U.S.-F	Japan-M	Japan-F	U.K.-M	U.K.-F
MAE (10^{-3})							
LC	2.7922	3.4509	2.1602	**1.5443**	1.8407	6.0328	1.7239
CBD	2.7055	3.5654	2.1269	1.8063	1.6097	5.5004	1.6242
RH	4.3511	4.6940	**1.0507**	10.2249	4.6734	4.1846	1.2790
M6	7.7354	5.0755	4.1334	13.7443	13.2380	**3.2548**	6.9667
M7	4.1549	6.5469	2.1567	3.4695	2.2692	7.8506	2.6364
JS	**2.4431**	**3.3166**	2.5073	2.4449	**1.2392**	3.9250	**1.2255**
MAPE (%)							
LC	30.06	19.86	19.56	27.20	51.27	33.70	28.75
CBD	31.31	24.49	21.92	33.35	56.09	35.12	16.92
RH	35.50	48.95	**15.48**	45.20	49.49	37.56	16.35
M6	62.65	66.77	28.46	72.01	106.47	59.51	42.69
M7	53.46	75.01	28.40	31.93	48.63	113.77	23.03
JS	**23.40**	**18.69**	18.84	**26.05**	**38.44**	**24.24**	**14.12**
RMSE (10^{-3})							
LC	3.4304	4.2548	3.0193	**1.5200**	2.3425	7.5688	1.8767
CBD	3.4474	4.1043	3.5449	2.3407	1.7703	6.7075	2.2169
RH	5.6247	4.4242	**1.5162**	14.4515	6.6458	4.6884	2.0223
M6	11.1262	4.6259	5.8970	21.3934	21.6742	**3.1600**	10.0066
M7	5.4530	8.0236	2.4172	5.6493	3.1144	9.5849	3.9288
JS	**2.9714**	**4.0266**	3.4556	2.7764	**1.4990**	4.4446	**1.6262**

Table 4.4: Forecasting performance metrics for forecasting 30 years

- The box plots of MAPEs from the JS model has the lowest median for all three forecasting years in Figure 4.4, which means the JS model produces the least absolute percentage prediction error compared to the other mortality models. Furthermore, the box plot of MAPEs for the JS model does not overlap with those for the famous LC and CBD models when forecasting 10 years; therefore, the JS model has much better performances than the LC and CBD models based on the box plots. Finally, the JS model has no extreme outliers when forecasting 10 and 20 years, while LC and CBD methods have big MAPE outliers. When forecasting 30 years, the JS model has the smallest interquartile ranges (IQR) and lowest median MAPE compared to all the other methods. Thus, it can be concluded that the JS model has the lowest forecasting error and is the best forecasting model among all selected mortality models.

- In Figure 4.5, the magenta $MAPE_t$ line for the JS model stays at the bottom of the figures for both genders of U.K. and Japan when we forecast 10 years, which means the

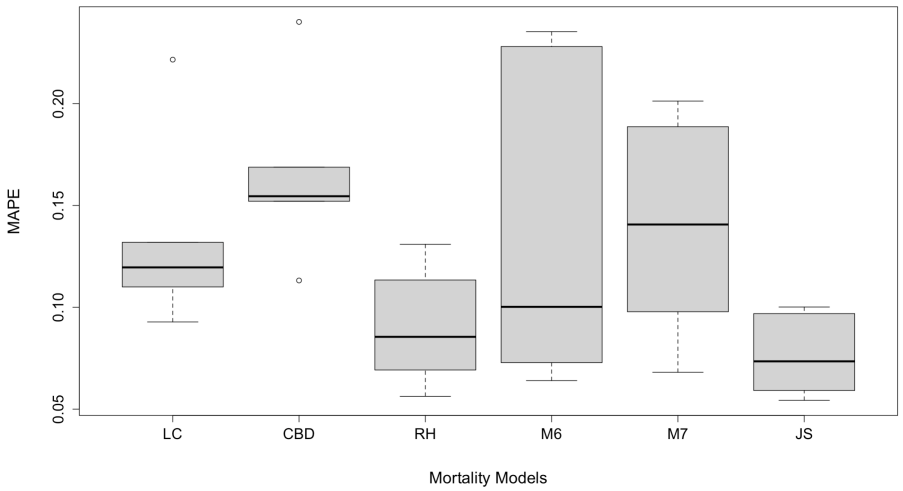
JS model leads to the lowest absolute percentage error compared to the other methods for these four populations. On the other hand, the RH and M6 models are the best forecasting models for males and females of U.S., respectively, as their $MAPE_t$ lines stay at the bottom of the figures for most of 10 forecasting years.

- The JS model is the best forecasting model when we forecast 20 years because the magenta $MAPE_t$ lines for the JS model are constantly at the bottom of all sub-figures of in Figure 4.6 for all six populations. Finally, when we forecast 30 years, observation from Figure 4.7 shows that the JS model is the best model for all populations except U.S. females for which the RH model is the best one.

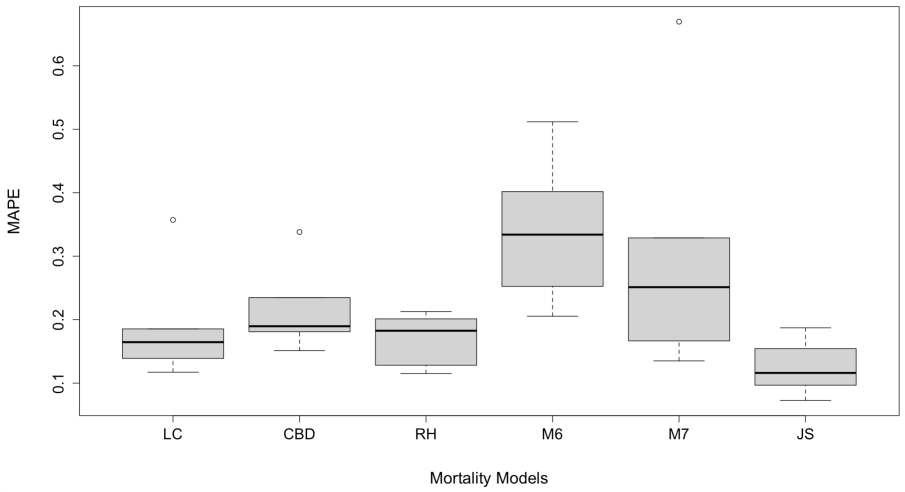
- It is interesting to note in Figures 4.5–4.7 that there is a sudden increase or decrease in $MAPE_{2020}$ for most of the mortality models when forecasting mortality rates for U.S. and U.K. populations, while $MAPE_{2020}$ stays around the trend of $MAPE_t$ for Japan populations. Figure 4.8 demonstrates that there is a sudden increase in $(1/60)\sum_{x=25}^{84}\ln(m_{x,\,2020})$ (the average of observed $\ln(m_{x,\,2020})$) for both genders of the U.S. and U.K., while the trends of $(1/60)\sum_{x=25}^{84}\ln(m_{x,\,t})$ remain stable for the Japan populations because Covid-19 pandemic that started in early 2020 increases the logarithm of central death rates for U.S. and U.K., but has not affected Japan's mortality trends. These sudden increases or decreases in $MAPE_{2020}$ can be explained as follows: the mortality models which overestimate the mortality rates in 2019 experience a sudden decrease in $MAPE_{2020}$, while the mortality models which underestimate the mortality rates in 2019 experience a sudden increase in $MAPE_{2020}$ as the Covid-19 pandemic raises the mortality rates in 2020.

4.4 Explore the Shrinkage effect of the James-Stein Estimation Method

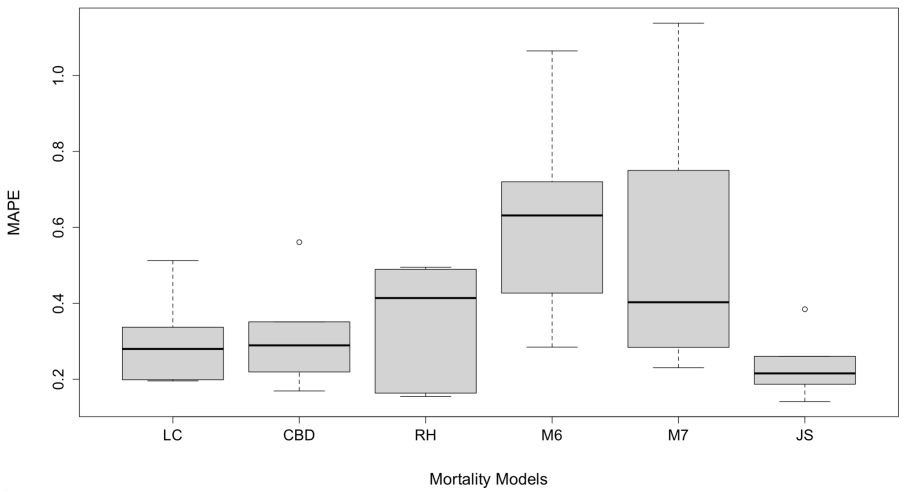
The James-Stein Estimation method is compared with the other mortality rate estimation methods in the previous section and achieves great results. The shrinkage behavior of the JS estimator is explored in this section through comparing the JS method estimation outputs with the vector of sample means (average of observed annual decrements), $\underline{\bar{\mathbf{Y}}}$. The baseball batting example in Figure 3.1 from Efron and Morris [11] shows that the baseball players are ordered according to their observed batting performances, and the observed batting averages are shrunk toward the JS estimation outputs as shrinkage effects of the JS estimation. Specifically, the best and worst performing players experience the most shrinkage effect towards the grant average (overall mean) of 0.265, Y_0, and the JS estimation method achieves lower estimation errors than the maximum likelihood estimation method due to this shrinkage behavior. Moreover, the best performing player is predicted to play worse than his historical batting average, while the worst performing player is predicted to



(a) Forecasting 10 years

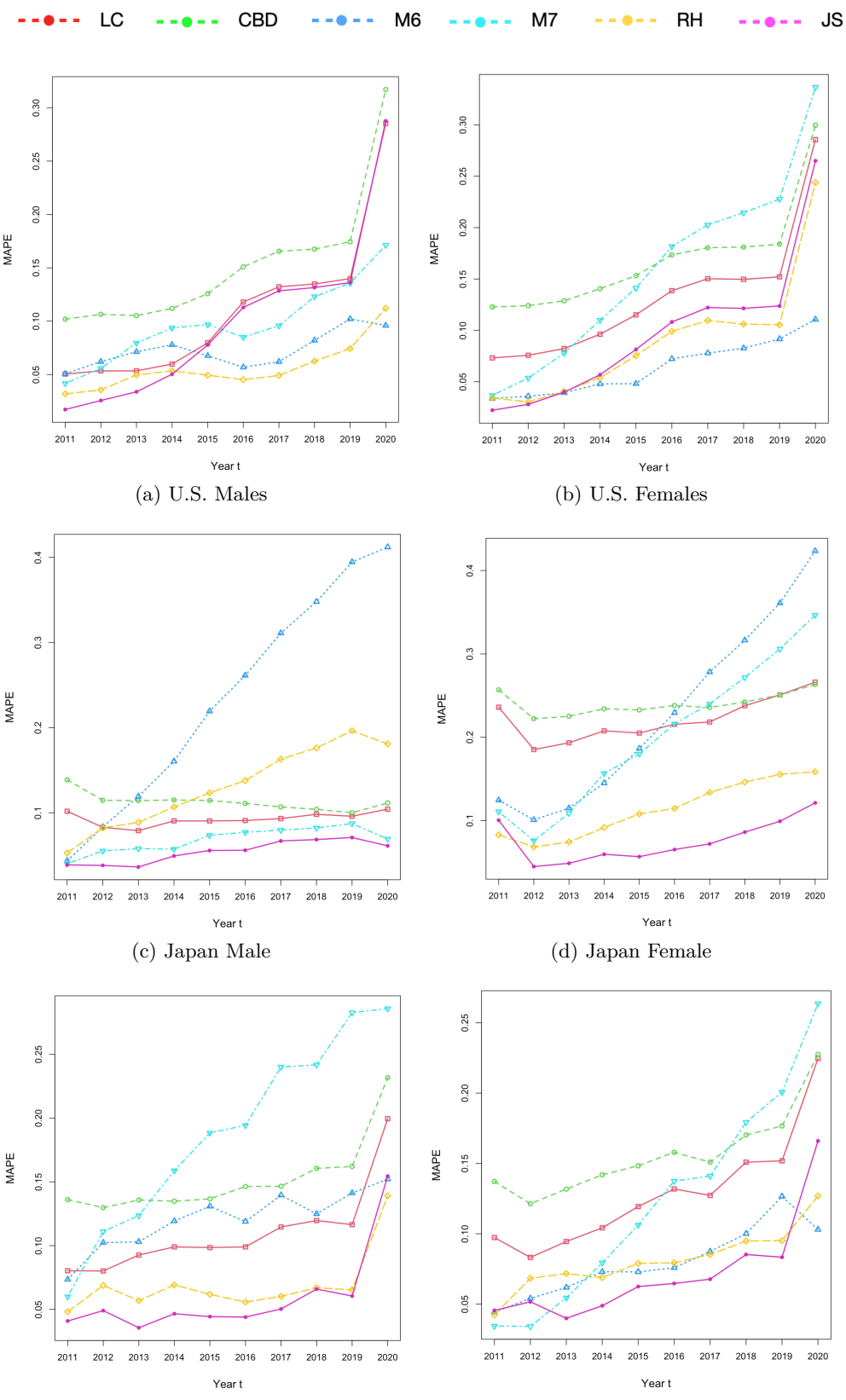


(b) Forecasting 20 years



(c) Forecasting 30 years

Figure 4.4: Box plots of MAPEs over all ages, years, and populations



(a) U.S. Males

(b) U.S. Females

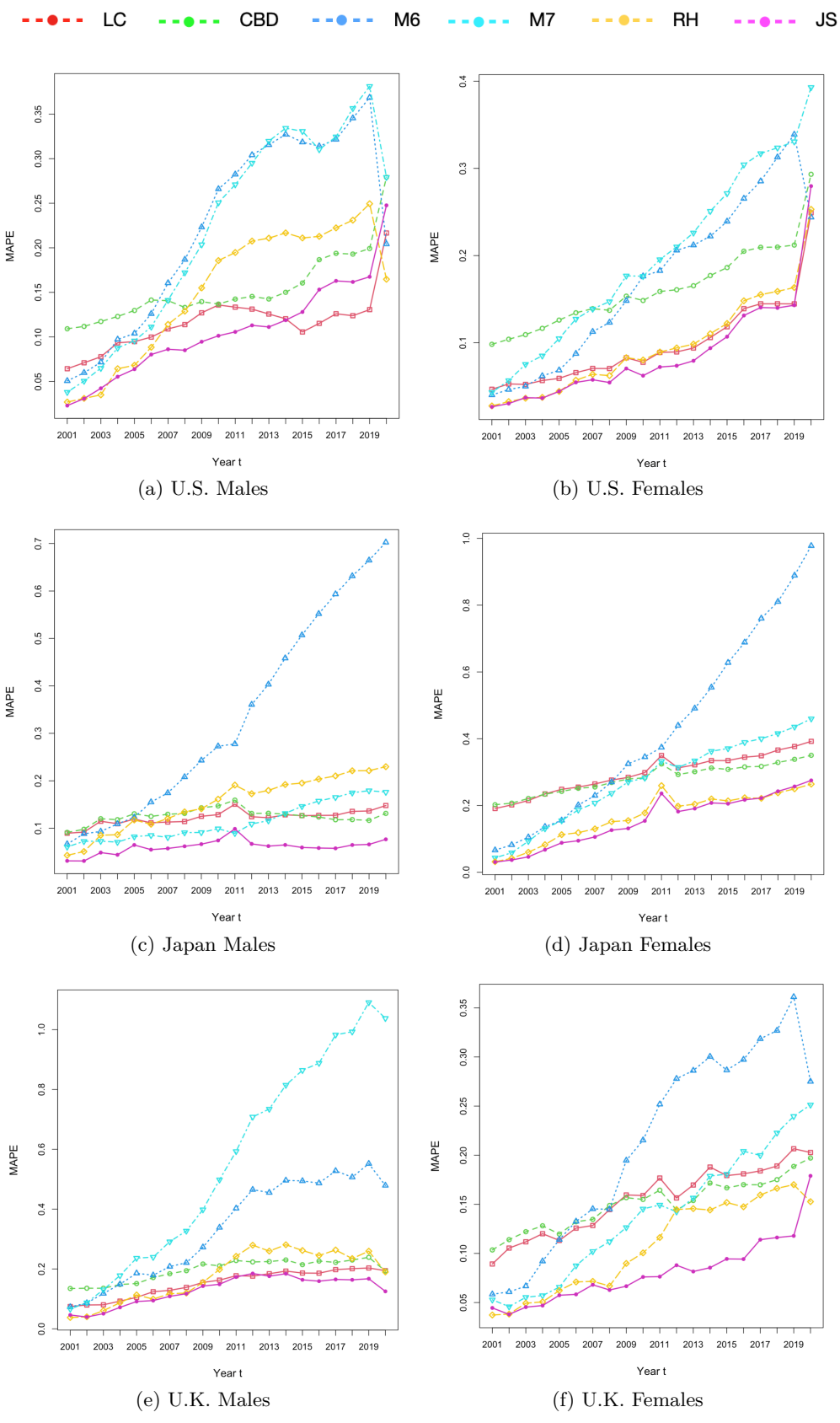
(c) Japan Male

(d) Japan Female

(e) U.K. Males

(f) U.K. Females

Figure 4.5: $MAPE_t$ against $t = 2011, \ldots, 2020$ (forecasting 10 years)



(a) U.S. Males

(b) U.S. Females

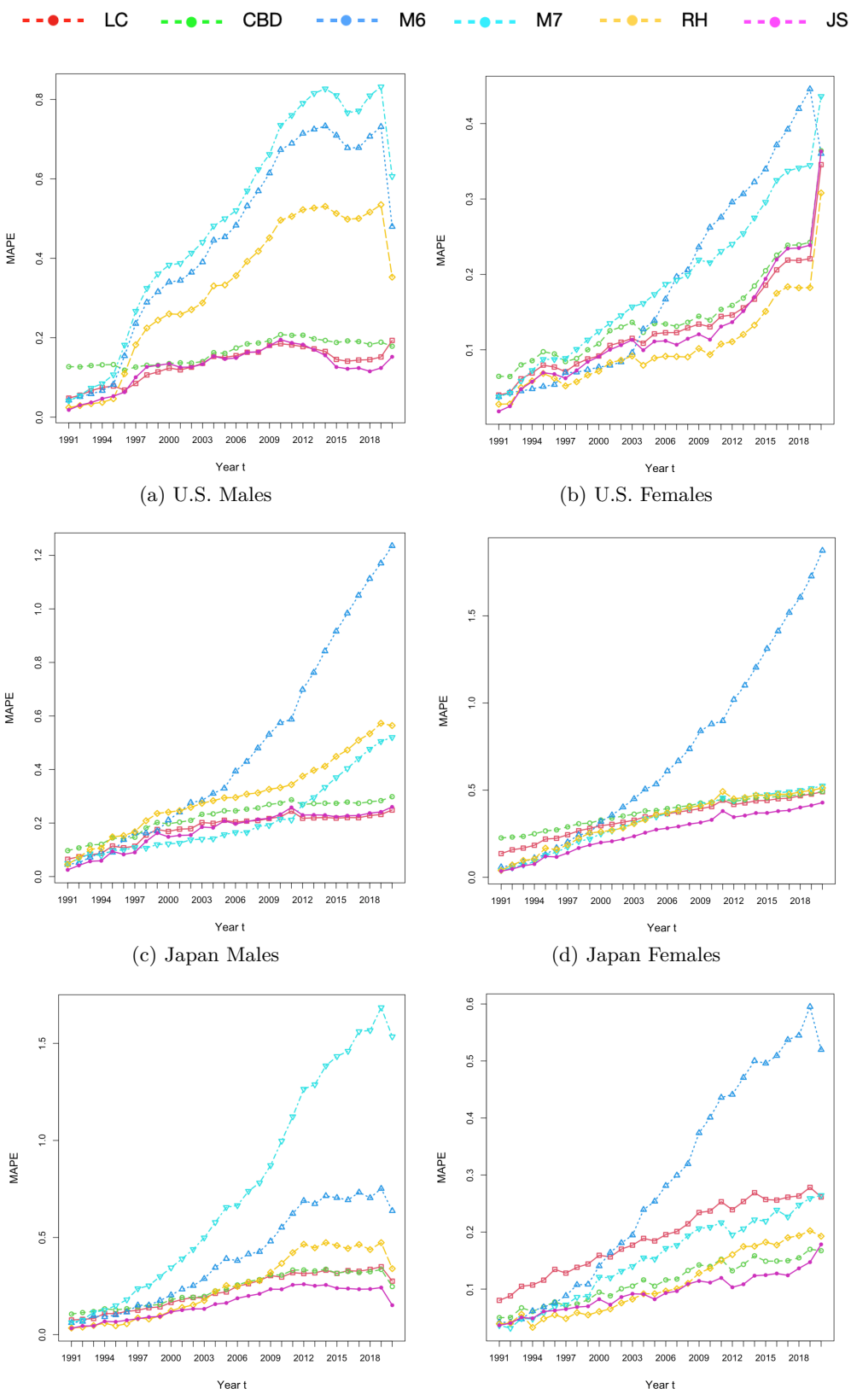
(c) Japan Males

(d) Japan Females

(e) U.K. Males

(f) U.K. Females

Figure 4.6: $MAPE_t$ against $t = 2001, \ldots, 2020$ (forecasting 20 years)



(a) U.S. Males

(b) U.S. Females

(c) Japan Males

(d) Japan Females

(e) U.K. Males

(f) U.K. Females

Figure 4.7: $MAPE_t$ against $t = 1991, \ldots, 2020$ (forecasting 30 years)

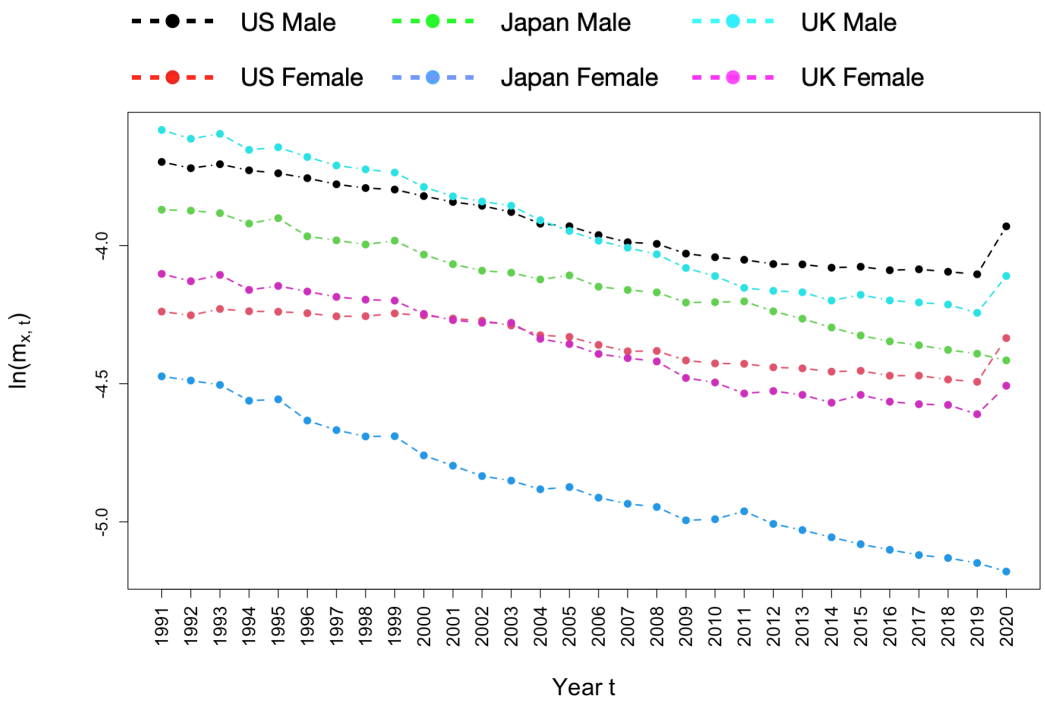


Figure 4.8: $(1/60)\sum_{x=25}^{84}\ln(m_{x,t})$ (average of observed $\ln(m_{x,t})$), $t = 1991, \ldots, 2020$

play better than his historical batting average. Thus, the example in Efron and Morris [11] demonstrates that the JS estimation method sacrifices forecasting accuracy at two extreme ends to achieve the overall forecasting performance. Therefore, it is interesting to analyze if the JS estimation method has the similar shrinkage behavior when we apply to the mortality rate estimation problem, and how does the modified James-Stein estimator shrinks the observed sample mean for all ages and populations.

An entry in the vector of averages of observed decrements, $\underline{\bar{\mathbf{Y}}}$, is the historical sample mean of annual mortality rate improvements for an age, and the observed overall mean, Y_0, is the overall mean of annual mortality rate improvements for all ages in a population. Figure 4.9 plots three curves of overall mean (black), ordered sample means (blue), and corresponding James-Stein estimates (red), based on 49 U.S. males annual decrements $Y_{x,t}$s for each age x from the $[25, 84] \times [1951, 2000]$ training data set, against the ordered sample means for all ages. It shows that all ages have mortality rate improvements, since all entries of $\underline{\bar{\mathbf{Y}}}$ are negative across all ages. Next, the JS estimated average decrements are also all negative, and the ordered sample means (blue) are shrunk towards the corresponding JS estimated means (red) across all ages. Specifically, the largest and smallest sample means (averages of annual decrements) at the two extreme ends experience the largest amount of shrinkage compared to the other sample means in the middle of the linear line. This observation matches with the JS shrinkage behavior in the baseball batting average example from Efron and Morris [11]. Because the modified JS estimation method uses the shrinkage

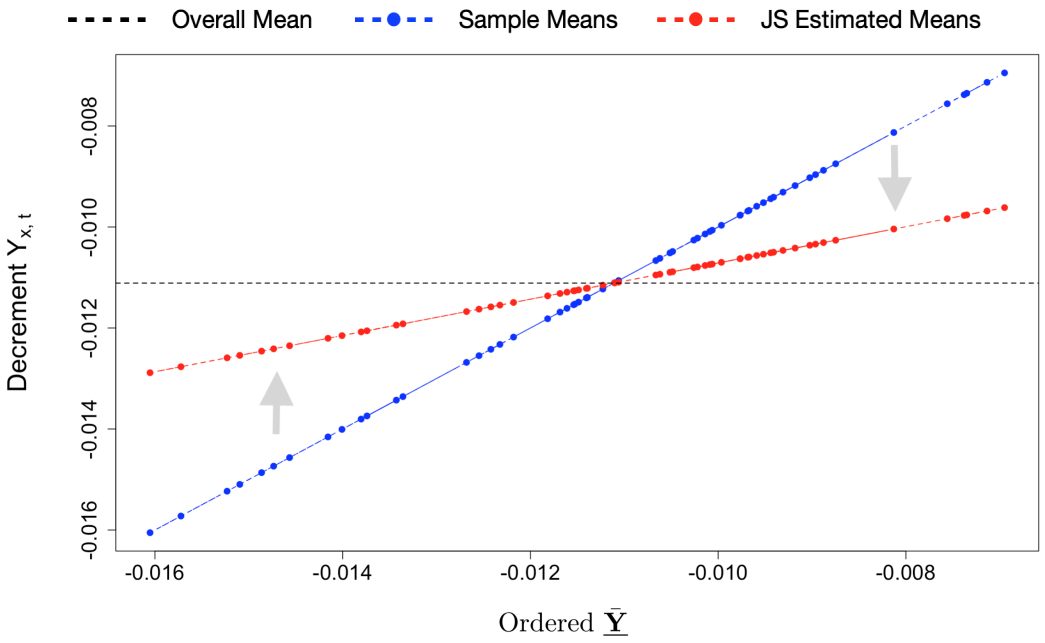


Figure 4.9: Comparison of overall mean (black), ordered sample means (blue), and corresponding James-Stein estimates (red) based on 49 U.S. males annual decrements $Y_{x,t}$s from the $[25, 84] \times [1951, 2000]$ training data set

factor $(1 - \hat{w})$ to get a weighted average of $\bar{\underline{\mathbf{Y}}}$ (the sample means) and $Y_0 \cdot \underline{\mathbf{1}}$ (the overall mean) from the observed annual decrements, it leads to that the red plot (for the James-Stein estimated means) falls between the blue plot (for the sample means) and the black plot (for the overall mean). Moreover, the shrinkage factor neutralizes the JS estimates such that the ages with higher historical mortality rate improvements (left side of Figure 4.9) are shrunk upward towards the observed overall mean in black dashed line, while the ages with lower historical mortality rate improvements (right side of Figure 4.9) are shrunk downward towards the observed overall mean.

Furthermore, Figures 4.10–4.12 display the absolute shrinkage between the observed sample means and the JS estimates on the left column, as well as the age-specific sample means for each age on the right column. The shrinkage for annual decrement from the observed sample means (averages of annual decrements) to the JS estimates for annual decrement (used for forecasting outside the age-year fitting window (training data window) is defined as follows:

$$\begin{aligned} \text{shrinkage} &= \text{JS estimates} - \text{observed sample means} \\ &= \left[(1 - \hat{w}) \cdot \bar{\underline{\mathbf{Y}}} + \hat{w} \cdot Y_0 \cdot \underline{\mathbf{1}}\right] - \bar{\underline{\mathbf{Y}}} = \hat{w} \cdot (Y_0 \cdot \underline{\mathbf{1}} - \bar{\underline{\mathbf{Y}}}). \end{aligned}$$

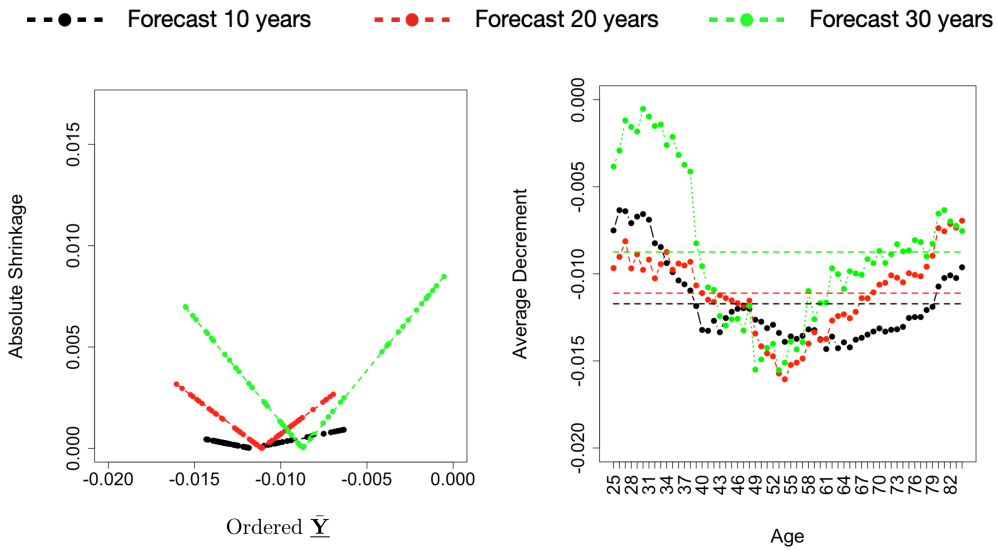
So, the absolute value of shrinkage for age x equals $\hat{w}$ times the difference between the sample mean for x and the overall mean. A sample mean far away from the overall mean

produces a large absolute value of shrinkage, and a sample mean close to the overall mean leads to a small absolute value of shrinkage.

The absolute values of shrinkage against the ordered sample means are plotted on the left column of Figure 4.10–4.12, and the vertices of the absolute value graphs are the overall means, Y_0, which are plotted as horizontal dashed lines in the sub-figures of the unordered sample means against age x on the right column of Figure 4.10–4.12. The sample means (average of annual decrements) that are larger (smaller) than Y_0 are shrunk downwards (upwards), and thus the sample means close to the two extreme ends experience large absolute values of shrinkage due to the James-Stein estimation process. Moreover, a wider forecasting year normally has a higher absolute value of shrinkage, such that the absolute value of shrinkage increases when moving from black dashed lines (for 10 years) to green dashed lines (for 30 years) across all three populations. This is because a longer forecasting window indicates a shorter fitting window (and thus, a smaller value of T) which leads to a higher value of $\hat{w}$ that places a more weight on the overall decrement mean, Y_0, than the observed sample means, $\underline{\bar{\mathbf{Y}}}$. Thus, the absolute value of shrinkage is high when the forecasting window is wide. Next, all three female populations in Figures 4.10–4.12 have higher historical mortality rate improvements than the male populations, since the horizontal lines (the overall decrement mean) for females are lower than those for males.

Furthermore, the width of the absolute value of shrinkage graphs for U.S. and Japan populations become wider when the forecasting window gets wider. For instance, Figure 4.10 (a) shows the green graph (for 30 years) is wider than the red (for 20 years) and black (for 10 years) graphs. However, the case of U.K. does not follow the same situation as that for the U.S. and Japan. Actually, the width of a graph for the ordered $\underline{\bar{\mathbf{Y}}}$ in sub-figures (a) and (c) is actually the vertical range of the unordered $\underline{\bar{\mathbf{Y}}}$ in sub-figures (b) and (d).

In conclusion, the JS model achieves better results than the other mortality models because it takes account of both the average annual decrement for each age and the overall annual decrement for all ages. In this mortality rate application problem using the modified JS estimator, by observing Figure 4.9, the age with small (see a blue bullet in the upper-right part) historical annual mortality rate improvement is moved down and predicted to have a bigger (see the corresponding red bullet in the upper-right part) future annual mortality rate improvement, whereas the age with big (see a blue bullet in the lower-left part) historical annual mortality rate improvement is moved up and predicted to have a smaller (see the corresponding red bullet in the lower-left part) future annual mortality rate improvement. Therefore, the JS estimator is indeed a biased estimator as indicated in Efron and Morris [11], such that it sacrifices a bit of accuracy at the ages with extremely high or low historical annual mortality rate improvements to achieve a better overall prediction performance.



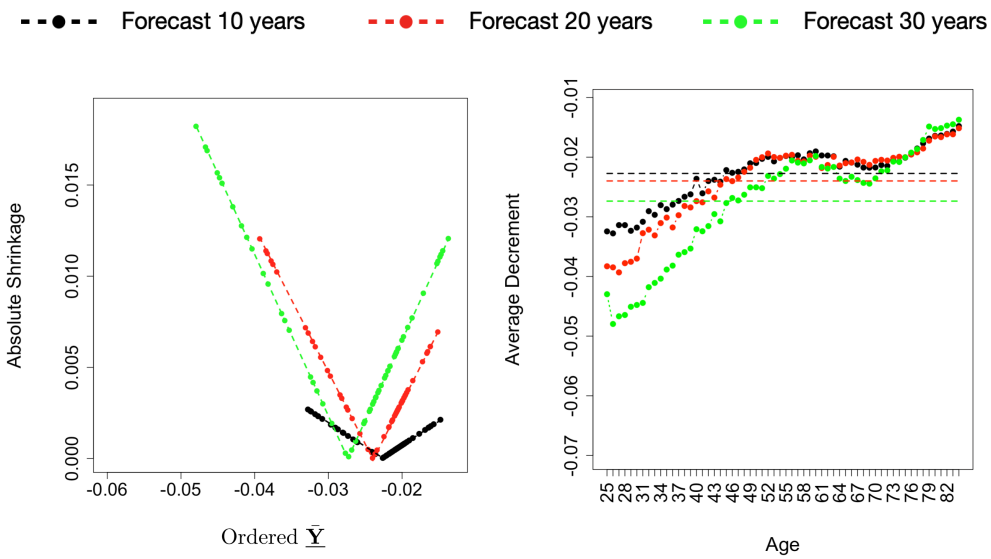
(a) U.S. Males

(b) U.S. Males

(c) U.S. Females

(d) U.S. Females

Figure 4.10: (a) and (c): absolute value of shrinkage against ordered $\underline{\bar{\mathbf{Y}}}$. (b) and (d): unordered $\underline{\bar{\mathbf{Y}}}$ against age x for the U.S.



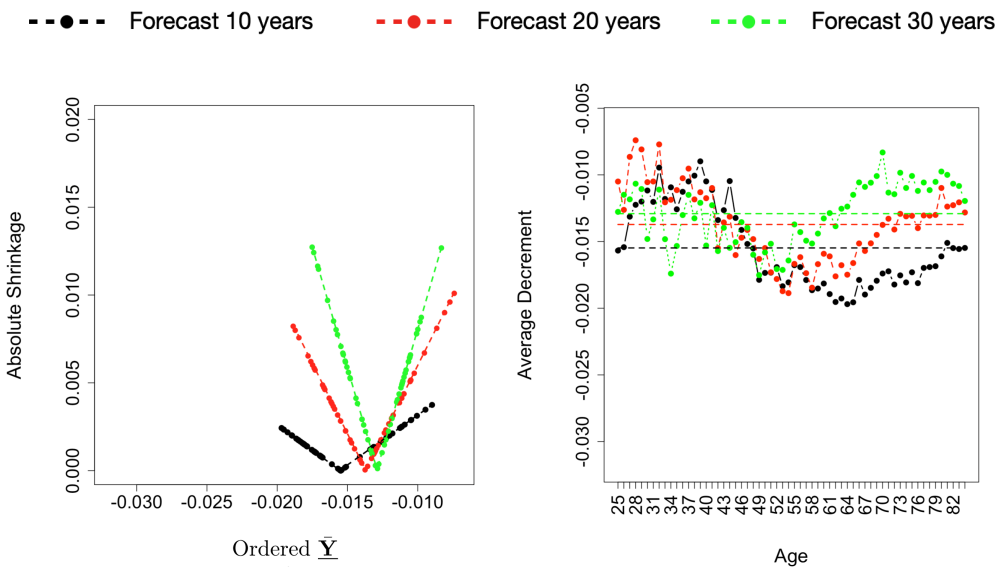
(a) Japan Males

(b) Japan Males

(c) Japan Females

(d) Japan Females

Figure 4.11: (a) and (c): absolute value of shrinkage against ordered $\underline{\bar{\mathbf{Y}}}$. (b) and (d): unordered $\underline{\bar{\mathbf{Y}}}$ against age x for Japan



(a) U.K. Males

(b) U.K. Males

(c) U.K. Females

(d) U.K. Females

Figure 4.12: (a) and (c): absolute value of shrinkage against ordered $\underline{\bar{\mathbf{Y}}}$. (b) and (d): unordered $\underline{\bar{\mathbf{Y}}}$ against age x for the U.K.

Chapter 5

Conclusions

The James-Stein estimation method shocks the statistical world with its ability to achieve a better prediction than the maximum likelihood estimation when predicting under the multivariate environment. In order to apply the JS method to the mortality rate forecast problem, the original JS estimator, inspired by the credibility theory, is modified to pool information from each individual mean and the overall mean to achieve a weighted average of these two means by using a shrinkage estimator.

The JS estimation method is compared with five mortality models including Lee-Carter (LC) model, Cairns-Blake-Dowd (CBD) model, M6 and M7 models (two extensions of the CBD model), and Renshaw-Haberman (RH) model. Historical mortality data for the U.S., the U.K., and Japan are used for the numerical illustrations and comparisons. The JS model outperforms all other models and achieves the lowest average errors across all three performance metrics, including mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) when predicting mortality rates for 10, 20, and 30 years. Moreover, the JS model predicts the mortality rates exceptionally well for U.K. males when forecasting 10 years, females of all three countries when forecasting 20 years, and U.S. males, Japan females and U.K. females when forecasting 30 years, where the JS model leads to the lowest error at each of three performance metric for those populations. Therefore, although the JS model is not the best mortality model for each single population, it ultimately leads to the lowest average error over all six populations. Thus, the James-Stein estimator sacrifices the accuracy at some ages and in some populations to achieve a better overall prediction performance than the other five mortality models.

Finally, the shrinkage effect of the JS estimation method is explored in the last section of the previous chapter. The JS model achieves a better estimation performance than other mortality models by shrinking the ages with extremely high or low average of annual mortality rate improvements toward the overall annual mortality rate improvement. One thing to note is that the JS estimation method has not replaced the MLE method in most of applications due to a major drawback mentioned in Efron [8] that the JS method results

in a high prediction error with a genuine outlier prediction. For instance, the score for an excellent player is shrunk drastically towards the average score for all players under the JS estimation method in the baseball player example from Efron [8], which leads to a high prediction error compared to the MLE method. Therefore, the JS method cannot handle the outlier cases very well, and it uses the accuracy for the outlier cases as trade-offs to achieve a better overall accuracy across different variables. In the example for numerical illustration, the JS model tends to shrink the ages with high annual mortality rate improvements downwards and shrink the ages with low annual mortality rate improvements upwards, which leads to biases at some ages. However, it is rare that a very different annual mortality rate improvement for an age is treated as an outlier case. Thus, application of the JS estimation method to the mortality estimation problem has no of this drawback.

Therefore, although the James-Stein estimation method obtains a better prediction performance than the other widely used mortality models, it needs to be improved to have a better prediction without shrinking too much at the both ends of the averages of annual mortality rate improvements. For future studies, regularization factors and Bühlmann credibility analysis can be used to perform penalty-based shrinkage to regularize the size of shrinkage to prevent over-shrinkage at ages with extremely high or low annual mortality rate improvements.

Bibliography

[1] M.-P. Bergeron-Boucher and S. Kjærgaard. Mortality forecasting at age 65 and above: An age-specific evaluation of the Lee-Carter model. *Scandinavian Actuarial Journal*, 2022(1):64–79, 2021.

[2] A. Bozikas, I. Badounas, and G. Pitselis. Pricing longevity bonds under a credibility framework with limited available data. *Risks*, 10(5):96, 2022.

[3] J. D. Braverman. Credibility theory: A probabilistic development. *The Journal of Risk and Insurance*, 35(3):411–423, 1968.

[4] H. Bühlmann. Experience rating and credibility. *The Journal of the IAA*, 4(3):199–207, 1967.

[5] A. J. G. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718, 2006.

[6] A. J. G. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich. A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13(1):1–35, 2009.

[7] L. R. Carter and R. D. Lee. Modeling and forecasting us sex differentials in mortality. *International Journal of Forecasting*, 8(3):393–411, 1992.

[8] B. Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press, Cambridge, 2010.

[9] B. Efron and T. Hastie. *James–Stein Estimation and Ridge Regression*, page 91–107. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2016.

[10] B. Efron and C. Morris. Families of minimax estimators of the mean of a multivariate normal distribution. *The Annals of Statistics*, 4(1):11–21, 1976.

[11] B. Efron and C. Morris. Stein's paradox in statistics. *Scientific American*, 236(5):119–127, 1977.

[12] J. Hausser and K. Strimmer. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *Journal of Machine Learning Research*, 10:1469–1484, 2008.

[13] W. James and C. Stein. Estimation with quadratic loss. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 361–379, Berkeley, California, 1961.

[14] W. James and C. Stein. *Estimation with quadratic loss*, page 443–460. Springer Series in Statistics. Springer, 1992.

[15] P. Jorion. Bayes-Stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279–292, 1986.

[16] R. J. Marshall. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics*, 40(2):283–294, 1991.

[17] A. E. Renshaw and S. Haberman. A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570, 2006.

[18] R. J. Rossi. *Likelihood-based Estimation*, pages 223–274. John Wiley Sons, Inc., 2018.

[19] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 197–206, Berkeley, California, 1956.

[20] C. C.-L. Tsai and T. Lin. A Bühlmann credibility approach to modeling mortality rates. *North American Actuarial Journal*, 21(2):204–227, 2017.

[21] C. C.-L. Tsai and T. Lin. Incorporating the Bühlmann credibility into mortality models to improve forecasting performances. *Scandinavian Actuarial Journal*, 2017(5):419–440, 2017.

[22] C. C.-L. Tsai and S. Yang. A linear regression approach to modeling mortality rates of different forms. *North American Actuarial Journal*, 19(1):1–23, 2015.

Appendix A

Mortality Models

A.1 Lee-Carter Model

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{A.1}$$

for $x = x_L, \ldots, x_U$ and $t = t_L, \ldots, t_U$, where

- $m_{x,t}$ is the central death rate for an individual aged x in year t;
- α_x is the average age-specific mortality;
- κ_t is the time trend factor in year t;
- β_x is the age-specific reaction to the time trend factor; and
- $\epsilon_{x,t}$, $t = t_L, \ldots, t_U$, are the i.i.d. residual errors that are not captured by the model.

Two constraints:

- $\sum_{x=x_L}^{x_U} \beta_x = 1$, and
- $\sum_{t=t_L}^{t_U} \kappa_t = 0$.

A.2 Renshaw-Haberman Model

$$\ln(m_{x,t}) = \alpha_x + \beta_x^{(1)} \kappa_t + \beta_x^{(0)} \gamma_{t-x} + \epsilon_{x,t}, \tag{A.2}$$

for $x = x_L, \ldots, x_U$ and $t = t_L, \ldots, t_U$, where

- $m_{x,t}$ is the central death rate for an individual at x age and t year;
- α_x is the average age-specific mortality;
- κ_t is the time trend factor in year t;
- $\beta_x^{(1)}$ is the age-specific reaction to the time trend factor;
- γ_{t-x} is the age-year trend cohort factor;
- $\beta_x^{(0)}$ is the age profile change in reaction to the age-time trend cohort factor; and
- $\epsilon_{x,t}$, $t = t_L, \ldots, t_U$, are the i.i.d. residual errors that are not captured by the model.

Four constraints:

- $\sum_{x=x_L}^{x_U} \beta_x^{(1)} = 1$,
- $\sum_{x=x_L}^{x_U} \beta_x^{(0)} = 1$,
- $\sum_{t=t_L}^{t_U} \kappa_t = 0$, and
- $\sum_{c \in C} \gamma_c = 0$, where $C = \{t - x : t = t_L, \ldots, t_U; x = x_L, \ldots, x_U\}$.

A.3 Cairns-Blake-Dowd Model

$$\ln\left(\frac{q_{x,t}}{1 - q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}\,(x - \bar{x}) + \epsilon_{x,t}, \tag{A.3}$$

for $x = x_L, \ldots, x_U$ and $t = t_L, \ldots, t_U$, where

- $q_{x,t}$ is the probability that an individual aged x in year t will die within one year;
- $\bar{x}$ is the average age over $x = x_L, \ldots, x_U$;
- $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are the time trend factors in year t; and
- $\epsilon_{x,t}$, $t = t_L, \ldots, t_U$, are the i.i.d. residual errors that are not captured by the model.

No constraint.

A.4 CBD Model with a Cohort Effect Term (M6)

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}\,(x-\bar{x}) + \gamma_{t-x} + \epsilon_{x,t}, \tag{A.4}$$

for $x = x_L, \ldots, x_U$ and $t = t_L, \ldots, t_U$, where

- $q_{x,t}$ is the probability that an individual aged x in year t will die within one year;
- $\bar{x}$ is the average age over $x = x_L, \ldots, x_U$;
- $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are the time trend factors in year t;
- γ_{t-x} is the age-year cohort trend factor; and
- $\epsilon_{x,t}$, $t = t_L, \ldots, t_U$, are the i.i.d. residual errors that are not captured by the model.

Two constraints:

- $\sum_{c \in C} \gamma_c = 0$, and
- $\sum_{c \in C} c\,\gamma_c = 0$, where $C = \{t - x : t = t_L, \ldots, t_U; x = x_L, \ldots, x_U\}$.

A.5 CBD Model with Cohort Effect and Quadratic Terms (M7)

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}\,(x-\bar{x}) + \kappa_t^{(3)}\,((x-\bar{x})^2 - s_x^2) + \gamma_{t-x} + \epsilon_{x,t}, \tag{A.5}$$

for $x = x_L, \ldots, x_U$ and $t = t_L, \ldots, t_U$, where

- $q_{x,t}$ is the probability that an individual aged x in year t will die within one year;
- $\bar{x}$ is the average age over $x = x_L, \ldots, x_U$;
- $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, and $\kappa_t^{(3)}$ are the time trend factors in year t;
- s_x^2 is the average value of $(x-\bar{x})^2$ over $x = x_L, \ldots, x_U$; and
- γ_{t-x} is the age-year cohort trend factor; and
- $\epsilon_{x,t}$, $t = t_L, \ldots, t_U$, are the i.i.d. residual errors that are not captured by the model.

Three constraints:

- $\sum_{c \in C} \gamma_c = 0$,
- $\sum_{c \in C} c\,\gamma_c = 0$, and
- $\sum_{c \in C} c^2\,\gamma_c = 0$, where $C = \{t - x : t = t_L, \ldots, t_U; x = x_L, \ldots, x_U\}$.